The Citation Culture

Paul Wouters

Doctoral Thesis -
University of Amsterdam

Next ...

To Contents Page ....

The Citation Culture

ACADEMISCH PROEFSCHRIFT
ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam,
op gezag van de Rector Magnificus
prof. dr. J. J. M. Franse
ten overstaan van een door het college
voor promoties ingestelde commissie
in het openbaar te verdedigen
in de Aula der Universiteit
op dinsdag 9 maart 1999, te 15.00 uur
door

Paulus Franciscus Wouters
geboren te Amsterdam

Voor Thomas en Marije
In herinnering aan Jan

Promotor: prof. dr. S. S. Blume
Co-promotor: dr. R. P. Hagendijk
Faculteit der Scheikunde

©Paul Wouters

Select Chapter Title from Contents
Listing below and Click

Contents

List of Figures

List of Tables

Voorwoord

Het onderzoek dat aan dit proefschrift ten grondslag ligt is mogelijk gemaakt door de belangeloze ondersteuning van bijzonder veel mensen. Dankzij hun inzet kon ik putten uit een grote verscheidenheid van bronnen en ervaringen. Ik ben hen daarvoor bijzonder erkentelijk en ik hoop dat zij enig plezier kunnen beleven aan het resultaat.

Het netwerk waarin dit proefschrift is ingebed is uiteraard heterogeen van aard en bevat zowel formele als informele elementen. Het private deel is het belangrijkste en voor mij staat Dik daarin als levenspartner en sparring partner centraal. Zonder zijn warm tegenvuur, intelligent commentaar, en geestelijk voedsel zou dit boek er eenvoudigweg niet zijn.

In het publieke netwerk neemt de vakgroep Wetenschaps- en Technologiedynamica een centrale plaats in. Ik heb mijn vakgroep ervaren als een stimulerende omgeving, niet het minst omdat ze intellectuele nieuwsgierigheid paart aan variëteit van karakter. Rob Hagendijk heeft mijn verwachtingen over begeleiding vér overtroffen. Ik had niet gedacht dat in deze overspannen academische tijden nog zo'n warmte en kritische aandacht in de begeleiding van een promovendus zou worden geïnvesteerd. Ook van het commentaar van Stuart Blume heb ik genoten. Van de wijze waarop hij in zijn schrijven met twijfel omgaat heb ik, komende uit een polemische politieke omgeving (de CPN), veel opgestoken.

Een bijzondere bijdrage aan dit proefschrift is geleverd door Lyuba Gurjeva die haar doctoraalscriptie heeft gewijd aan de sciëntometrie in Rusland. Hoofdstuk 4 is in belangrijke mate ook haar hoofdstuk.

Ook alle andere, vroegere en huidige, leden van de vakgroep hebben direct of indirect de loop van dit onderzoek met hun onophoudelijk kritische commentaar beïnvloed waarvoor ik hen allen hartelijk dank. Bijzondere vermelding verdient Loet Leydesdorff die me als senior-onderzoeker en co-auteur het handwerk van het (kwantitatieve) onderzoek leerde en een constante intellectuele prikkeling was. Stimulerende indrukken bleken niet beperkt tot Amsterdam: vooral Trudy Dehue maakte in het begin van de voor mij hernieuwde academische loopbaan indruk en zette daarmee een stevig stempel. Van mijn geleerde en zeer geleerde collega's van de onderzoekschool waren ook Wiebe Bijker, Hans Harbers, Annemarie Mol, Arie Rip en Gerard de Vries altijd bereid me op weg te helpen.

Ik hoop dat mijn mede-promovendi van AIO-netwerk, onderzoekschool en vakgroep die zich over mijn halfbakken producten hebben gebogen niet al te zeer teleurgesteld zijn over dit boek-in-wording dat uiteindelijk ook door hen is gevormd, in het bijzonder door Marc Berg, Ruth Benschop, Adrienne van den Boogaard, Gertrud Blauwhof, Carla van El, Patricia Faasse, Willem Halffman, Ruud Hendriks, Jessica Mesman, Annemiek Nelis, Bernike Pasveer, Irma van der Ploeg, Floor Rikken, Kaat Schulte-Fischedick, Frank Wamelink. Apart vermelding verdienen mijn kamergenoot Ad Prins en Anne Beaulieu met wie ik verrassend veel heb kunnen delen.

Daphne Visser-Lees dank ik hartelijk voor haar nauwgezette redactie van de Engelse tekst. Ik wil ook de huidige en vroegere leden van het secretariaat en de administratie bedanken voor hun inzet, evenals de medewerkers van de bibliotheek. Thomas Wouters en Sylvan Katz ben ik erkentelijk voor hun advies met betrekking tot programmatuur.

Dit onderzoek is deels gebaseerd op archiefonderzoek. Eugene Garfield stelde zijn privé-archief onvoorwaardelijk ter beschikking. Ik ben hem en zijn staf daarvoor erkentelijk. Ook Arie van Heeringen en de RAWB-staf waren zo gastvrij en tolereerden me wekenlang op hun zolder. Voor toegang tot archieven ben ik ook dank verschuldigd aan J. Merton England (NSF), de Stichting FOM, de staf van La Villette, het Ministerie van OC& W, Ben Martin, Francis Narin, Tibor Braun, het CWTS in Leiden, en Hildrun Kretschmer. De onderzoekschool WTMC, NWO en ASIS hebben reisbeurzen voor dit onderzoek verstrekt.

Beverly Bartolomeo, Donald D. de Beaver, Manfred Bonitz, Tibor Braun, Emiel Broesterhuizen, Michel Callon, Stephen Cole, Jean-Pierre Courtial, Bob Coward, Suzan Cozzens, Leo Egghe, Helen Gee, Eugene Garfield, Michael Gibbons, Wolfgang Glänzel, Isabelle Gomez en haar collega's, Arie van Heeringen, Diana Hicks, Wim Hutter, Phoebe Isard, Sheila Jasanoff, Sylvan Katz, Mike Koenig, Hildrun Kretschmer, Bruno Latour, Joshua Lederberg, Cees le Pair, Terttu Luukkonen, Morton Malin, Ben Martin, Robert King Merton, Henk Moed, Francis Narin, Ton Nederhof, Ton van Raan, Henk Rigter, Arie Rip, Jo Ritzen, Ronald Rousseau, András Schubert, Irving Sher, Len Simon, Henry Small, Jan van Steen, Peter Tindemans, William Turner, John Ziman, en Harriet Zuckerman waren zonder uitzondering bereid mijn ongetwijfeld vaak domme vragen te beantwoorden, waarvoor dank.

Ik ben de deelnemers aan de nationale en internationale conferenties waar ik mijn tussenproducten presenteerde erkentelijk voor hun aandacht en stimulans, alsmede de referees van de tijdschriftartikelen[1] Dit geldt in het bijzonder Diana Hicks, Sylvan Katz, Henry Small, Henk Moed en Ton van Raan, die nooit te beroerd bleken zich te laten provoceren tot uitdagende debatten.

Tijdens dit onderzoek kon ik profiteren van de wijsheid uit twee werelden, de academische en de journalistieke. Indirect hebben mijn journalistieke collega's dan ook meer invloed op dit proefschrift uitgeoefend dan ze zich zullen realiseren. Simon Rozendaal rondde met zijn provocerende commentaar en warme ondersteuning m'n opleiding af van pamflettist tot journalist, die begonnen was op de redactie van *De Waarheid*. Ook met Willem Schoonen was het altijd prettig samenwerken en ik ben blij dat we dat weer hebben hervat. Hein Meijers, Liesbeth van de Garde en de redactie van *Hypothese* gaven me menige gelegenheid over wetenschap te schrijven, en het *Science Channel* team verschaft me dagelijks een stimulerende virtuele omgeving.

Zo tegen het eind van dit voorwoord wordt het tijd mijn familie en vrienden een dikke zoen toe te werpen. Van hen wil ik in de eerste plaats Elly Baan bedanken voor de vele leuke jaren die we samen hebben doorgebracht. Otto Middelkoop en Kees Hulsman voor de bijzondere vriendschap. Mijn ouders Jan en Bep (jij geeft het wel door hè, Bep?), en mijn zussen en broer Marja, Marc, Caroline en Debbie voor de warmte waarmee ze me van jongs af aan omgaven. En tot slot Thomas en Marije voor hun liefde en de lesjes die ze me gelukkig altijd nog leren.

Amsterdam, januari 1999

[1] Dit onderzoek heeft geresulteerd in de volgende tijdschriftpublicaties: Wouters (1992*a*), Wouters (1992*b*), Wouters & Leydesdorff (1994), Wouters (1997*a*), Wouters (1998*b*), Wouters (1999*a*), en Wouters (1999*b*).

Back to Contents Page

Chapter 1

Introduction

1.1 Introduction

The need for greater accountability of scientific researchers has created a number of new professions. The scientometrician is one of these experts. They measure science scientifically, often on behalf of science policy officals. They are specialized in rating and mapping the sciences, the social sciences, and the humanities with the help of huge databases derived from the scientific literature. This is not the whole story, however. The scientometrician is not only a policy oriented professional, but also a social scientist. Scientometricians have a core journal, *Scientometrics*, jointly published by Elsevier Science and the Hungarian publishing house Akadémiai Kiadó. There is an international conference which takes place every two years, organized by their scientific association, the International Society for Scientometrics and Infometrics. Currently, there are a few hundred scientometricians in the world. They vary from a lone individual who is part of a research library or history of science department, to a large collective with around twenty full-time researchers.

The professional scientometrician emerged in the sixties. Their creation is intimately linked to the invention of the *Science Citation Index* (*SCI*) in Philadelphia (United States). To date, scientometricians cannot boast of many successes. They do not seem to have had a great impact on the science policy of most countries. One cannot acquire a university degree in scientometrics. Its practitioners have to cope with resistance from the scientific community and their results are not always welcomed. Moreover, while scientometricians have only a relatively short history, their prospects are in doubt. It is not clear whether the profession of scientometrics will survive the ongoing revolution in scientific communication (Wouters 1996*c*). Computer mediated communication is rapidly becoming the principal medium for publication and dissemination of professional and scientific results. In a few years every scientific journal will be obtainable via computer networks and databases (Wouters 1997*b*, Wouters 1996*a*). These changes may lead to a crucial shift in the characteristics of the unit of publication, the scientific article. Currently, it is uncertain how this will affect the measurement of science and the development of scientometrics. Since the scientific article is one of the key objects in scientometrics, these changes in scientific publishing may very well lead to the early death of this new profession in its present form.

This study is not a history aimed at describing the specialty in its various stages of development in a more or less "complete" way. It might be characterized as a footnote to the available history of the sociology of science, providing at most a historically and sociologically informed theoretical argument about one aspect of this history. Yet, strange as it may seem, this micro-history relates to interesting features of present-day science in general. I will argue that the development of scientometrics can best be understood if we analyze this field as both indicator and embodiment of a recently emerged subculture in science: *The Citation Culture*. This subculture has unwittingly and subtly changed core concepts of modern science such as scientific quality and influence. Because of the citation culture, *being cited* has profoundly changed its meaning over the last two decades, with a number of consequences for scientists. It has moreover contributed to the transformation of the very essence of science policy, notwithstanding scientometrics's apparent lack of outstanding successes. This study tries to explore the possible meaning of the citation culture for the systematic generation of knowledge. To reach this goal, this analysis does not start with big concepts like power, science or truth. Instead, it will begin from the most humble entity in scientific articles, often merely visible in small-print: the reference.

1.2 Citing cultures

Today, a scientific publication is easily recognized by its footnotes, endnotes and references to other scientific articles or books. This is one of the features which make scientific texts so different from a journalist's story or a novel. A scientist seems to be — at least in his professional life — an annoyingly precise person, whose claims are painstakingly documented. Not only do researchers describe their own work in minute detail (Latour & Woolgar 1986), they also conscientiously cite colleagues whose publications they have used. As is well known, this literary style has not always been the norm; it emerged only during the second half of the nineteenth century (Bazerman 1988). The present-day ensemble of norms, rules, practices and interpretations, which are invoked by researchers every time they cite someone's work, entertain complex relationships with one another. These norms and rules do not determine citing practices in the strict sense nor do they indicate the clear meaning of the reference. Norms may even contradict one another. At the same time, a researcher is not free to do as he pleases. He must be able to justify his citing action in terms of the norms and rules of his specialty. The rules do not exist independently of the actions, however. They exist "within" the citing actions while they are nevertheless different from them. They fulfill the role of a resource which both enables and constrains researchers in their citing. This type of relationship between structure and action, rule and behaviour, is typical of cultural phenomena in general[1]. Therefore, this

[1]Culture is an ambiguous concept. This study follows Goudsblom (1962,1970) and Hagendijk (1996). Culture includes not only the ensemble of ideas and patterns of behaviour in a certain society, but also the relationships between society and nature as a whole. This perspective entails no

study speaks of *the citing culture* in science.

Citing behaviour seems to vary according to personal traits. Whereas one author will devote detailed attention to the list of references, another could not be less interested (though this cannot be said too loudly). Nevertheless, the overall citing properties of the publications within a certain field share the same characteristics. The mathematician tends not to cite many publications. The biomedical researcher, on the other hand, is not afraid to cite hundreds of articles. The historian also likes references, but in a different way. The literary scholar goes about citing in quite another way. It seems therefore better to speak of the citing culture in the plural form. The sciences host many types of citing culture, each slightly different from the other. A conceptual core that is mutually shared by every one of them cannot be isolated; the various citing cultures resemble one another, as members of one family do. It is possible, of course, to abstract certain general notions and claim that these constitute the core. For example, a scientist is supposed to cite honestly: he must have read the article and have found it useful in some way. The question is, however, in what way this differs from the generally accepted norm of honesty. The moment one tries to become more concrete, and asks what it means to cite honestly and correctly, the answer becomes specialty-bound. Citing cultures not only differ between specialties, they also vary between journals. This is not exclusive to typographical format. It also has to do with the type of reference, its number, its position in the text etcetera. Thus, the historical development of scientific publishing since the nineteenth century has provided for a fairly stable ensemble of citing cultures in science.

1.3 Unintended consequences of being cited

The gradual development of regular citing behaviour in scientific publishing has created a new resource for research and policy: citation data. It did not take long before these data began to be used. With hindsight, it seems an almost inevitable outcome of some straightforward reasoning. If researchers cite the work they find useful, often cited ("highly cited") work is apparently more useful to scientists than work which receives hardly any citations at all. Hence, the number of times an article is cited, seems to be an accurate measure of its impact, influence or quality. The same is true of the collected articles of one particular scientist, research group, journal or even institution. The more they are cited, the greater their influence. Sloppy work will not often be cited, except in heated controversies — or so the reasoning goes. Therefore, citation frequency seems a good way of objectively measuring scientific usefulness, quality, or impact.

Whatever one's view on the import of being cited, citation frequency is generally supposed to measure something that already exists. This is based on an implicit realist perspective with respect to the process of scientific communication: the indicator is seen as a more or less direct upshot of scientists' activities. There-

"great divide" between culture and nature. A different definition of culture is given by Luhmann (1985, 224) according to which culture is the available supply of themes including their semantics that can be called upon in communication (Blom 1997, 141).

fore, citation analysis — the art of measuring numbers of citations — provides a window onto the communication processes between scientists. Consequently, scientometrics, in which citation analysis has a central position, is defined as the quantitative study of scientific communication (Narin 1976).[2]

This study questions these realist interpretations of measuring science by citations. It will be shown that the citation culture is not a simple aggregate or derivative of citing culture in science. The citation as used in scientometric analysis and science and technology indicators is *not* identical to the reference produced at the scientist's desk. This is the first claim of my study: the citation is the product of the citation indexer, not of the scientist. Citation analysis has only been feasible on a discernable scale since the invention of computerized citation indexes. This is also the reason that the *Science Citation Index* (*SCI*), the *Social Science Citation Index* (*SSCI*) and the *Citation Index for the Arts & Humanities* (*CI&H*) (all invented by the same man) are the dominant databases in citation analysis. Getting to know the citation a little better implies looking into the production of these indices. Therefore, this enquiry into the citation culture starts with the origin of its main component (chapter 2).

From the early years of this century, research librarians have systematically applied citation analysis (Gross & Gross 1927, Gross & Woodford 1931, Cole & Eales 1917, Broadus 1967, Brookes 1988, Cason & Lubotsky 1936, Earle & Vickery 1969, Raisig 1960, Fussler 1949, Burton 1959*a*, Burton 1959*b*, Barrett & Barrett 1957, Cole 1952, Dyson 1952). They collected data on the frequency with which journals were cited. Supposed to measure the usefulness of subscriptions to these journals for their clients the scientists, journal citation analysis was a tedious job, however, since lists of references of many articles in lots of different journals had to be collected to measure the citation frequency of even a single journal. This seems to have been the main reason for the relative scarcity of these citation analyses.

The situation changed abruptly, however, with the invention of the *Science Citation Index* by Eugene Garfield. Using the *SCI*, it took far less work to extract citation frequencies from the data. It became even possible to measure the frequency with which an individual was cited, a feat previously unheard of. Nevertheless, the scientific community was not enthusiastic. Many researchers did not even use the *SCI* for its stated purpose as a bibliographic tool — to find relevant publications in the exponentially growing mountain of scientific literature. Neither did many researchers use it to keep abreast of their citation status, a measure without clear meaning to many scientists. The prevailing reaction was hostile or indifferent. The difficult birth of the citation index relates, at least partly, to the translation process needed for the citation culture to prosper. The need for this translation process is the result of the novel way in which the *SCI* represents science.

[2]For a reflexive and constructivist systems-theoretical approach that also sees scientometrics as the study of scientific communications see Leydesdorff (1995).

1.4 An objective representation of science

1.4.1 Representation

In this study, the concept of representation is not taken to mean "mirroring reality". Scientometrics does not mirror science, neither does the scientific literature. In general, representing means both "speaking or acting on behalf of" and "being able to stand in for". Every representation is the product of the interaction between the phenomenon it represents and its own production rules[3]. Obviously, many types of representation exist. Knowledge, including scientific knowledge, can also be represented as a representation of the world[4]. A given body of knowledge is built upon other representations. Sometimes it makes sense to order these according to their contingency relations. Scientific literature for example is based on research and is one of its most important direct products. Relative to daily practice in laboratories, literature is therefore a "first order" representation. In the same vein, citation analysis and scientometrics are based on scientific literature and are another step removed from underlying research practice. In other words, they can be seen as "second order" representations of what goes on in laboratories. This study draws upon these two bodies of knowledge and practice and can therefore be seen as a "third order" analysis and representation. These different representations are related through translation, distortion and transformation, more than through linear reflection.

1.4.2 The *SCI*

The *Science Citation Index*[5] is not merely a bibliographic instrument. It also creates a new picture of science via bibliographic references found in scientific literature. As the *Terminology & Definitions* section of the *SCI* explains:

> The Citation Index is an alphabetic list of references given in bibliographies and footnotes of source articles arranged by first author. Each reference is followed by brief descriptions (citations) of the source articles which cite it.

In this way, the *SCI* provides a fundamentally new representation of science. There had been similar devices before. However, these were confined to certain disciplines; the *SCI* is the first citation index aimed at the whole of scientific literature. It creates an image of this type of literature in the same way as a telephone book creates an image of the inhabitants of a city.

---

[3]For the discussion of representation, knowledge and the politics of explanation in science and history see Bloor (1976), Ashmore (1989), Latour (1988), Woolgar (1988*b*), Woolgar (1988*a*), Lynch & Woolgar (1990), Hagendijk (1996), Huizinga (1937), Romein (1976*c*), Romein (1976*b*), Ankersmit (1990), Lorenz (1987, 1994), Tollebeek (1996), Luhmann (1992), and Maturana & Varela (1988).

[4]This view deviates from Ankersmit (1990) who pictures science in a surprisingly realist way. Contrary to his views, this study does not assume a fundamental divide between science and history.

[5]By the term *Science Citation Index* are also meant the *Social Science Citation Index* and the *Arts & Humanities Citation Index*, all published by the Institute for Scientific Information (ISI), except when otherwise indicated.

Scientific literature is a representation of scientific research, produced by selectively emphasizing some cognitive features and neglecting others (Knorr-Cetina 1981, Latour & Woolgar 1986). The *SCI* in its turn represents scientific literature (it does not use any elements of science outside this literature) and is, consequently, a second order representation of science. Every representation is different from its object. After all, without differences the representation would be pointless. The *SCI* creates these differences by the selection of features of the literature it processes. Since the resulting index structure cannot be made at will by its producers, they do not know beforehand what will result from their work. If we take "reality" to be that which resists (Hacking 1983, Latour 1984), scientific literature is the real, independently existing, object of the citation index[6]. Therefore, while the index depends on the literature, this relationship is not reciprocal, at least initially[7]. Because of this relationship between literature and index, the *SCI* can be perceived as an objective (i.e. non-subjective) representation of scientific literature: "When using citation data, we draw on a multi-disciplinary, objective, and internally consistent data base, the Science Citation Index" (Small & Griffith 1974).

Almost immediately after its first publication, the *SCI* data were used in citation analysis. This type of research claims to be objective due to the above mentioned objective character of the *SCI*: "Citation analysis is objective because it is based on written information that anyone can check. It is the aggregate of the subjective decisions of all publishing scientists" (Aaronson 1975). Given the massive amount of data contained in the *SCI*, advanced statistical techniques, like co-citation clustering, need to be used. However, this does not seem to diminish the objectivity of the analytical results:

> Many of the relationships we have uncovered are, of course, known to the specialists themselves, since they were established by their own citing patterns, but the perspective this method offers is far broader than can be achieved by any individual scientist. This is the crux of the method: the observed relationships are in substance those which have been established by the collective efforts and perceptions of the community of publishing scientists. Our task is to depict these relationships in ways that shed light on the structure of science. (Small & Griffith 1974)

Apparently, three points are important. First, the *SCI* portrays science from a nonobtrusive *outsider's* position. Therefore structures can be revealed which cannot be perceived in that form from the position of the researcher in the represented field. Second, scientists seem to get the citation pattern back that they produced themselves. This study will show that this is not as obvious as it seems. Third, making sense of the *SCI* requires specific procedures, except if one only

[6]This partly contradicts Lynch & Woolgar (1990, 13) who state: "our position is that representations and objects are inextricably interconnected". The fact that the human race is inextricably dependent on representations, does however not necessarily mean that every specific representation is inextricably interconnected with every object it represents. In this case study it is literature already in existence (a representation of science) which is processed by the indexers.

[7]There is feedback, though, which will be treated later in this chapter.

wishes to see one's own citation score. It is not very illuminating to read the index from the first page to the last. The patterns in the index can only be read with the help of statistical techniques. Far from diminishing the objectivity of citation analysis, these statistical manipulations of the data contribute to the validity of its results.

Not only are *SCI* and citation analysis engaged in an object–representation relationship to scientific literature (the basis of its perceived objectivity), the index is moreover applicable to the whole of scientific literature because it *neglects the substantive claims and counter-claims in the literature*. Whereas the scientific literature represents science by focusing on its cognitive claims (the content of the articles and books published), the *SCI* represents scientific literature by obliterating this content and focusing instead on its formal properties. It only processes references, author names, institutional addresses, titles, language names and types of publication[8]. This selection creates a new, unified representation of science, diverging from the compartmental picture one gets if one tries to read all scientific publications. This latter endeavour is not only impossible because of the vast number of journals and books published but because of the large number of different languages involved. Every specialty and discipline speaks its own language (de Wilde 1992)[9]. The *SCI* translates this tower of Babel into an integrated whole by drastically reducing its complexity.

This creates a host of new possibilities. For example, one specialty can be compared to another (Small & Griffith 1974). Moreover, as Garfield (1970) has it: "the *SCI* tells how each brick in the edifice of science is linked to all the others". Therefore, it is conceivable that maps of science can be created, an idea first put forward by geneticist Gordon Allen and later advocated by Derek de Solla Price.

> Such maps, it was hoped, can indicate the state of science in a particular year, and by their changes from year to year, the overall progress of science. (Small & Sweeney 1985*a*)[10]

In short, the *SCI* portrays science as a citation network. It is based on the assumption that no significant contributions to scientific knowledge are being missed in this way. Price (1965*a*) developed the following argument in the early years of citation analysis using the *SCI*:

> since 10 percent of all papers contain no references and another, presumbly almost independent, 10 percent of all papers are never cited, it follows that there is a lower bound of 1 percent of all papers on the number of papers that are totally disconnected in a pure citation network and could be found only by topical indexing or similar methods; this is a very small class, and probably a most unimportant one.

[8] This has varied somewhat over the years, but this does not affect the argument.

[9] This does not mean, of course, that this disciplinary structure of science would be static. New specialties are, on the contrary, constantly created at the interface of old ones.

[10] This was the foundation of ISI's project to produce Atlasses Of Science (Starchild et al. 1981, Garfield et al. 1984).

Given the regularity of its citing cultures, the representation of science as a citation network is generally seen as a reasonably accurate picture of science. This position common to scientometrics and the sociology of science is based on three assumptions:

1. The actual production of the citation index in Philadelphia does not fundamentally change the elements it uses. The *SCI* is consequently seen as the product, not of the indexers, but of the publishing and citing scientists.

2. The citing behaviour of scientists is assumed to be both sufficiently important and regular enough to shed light on the characteristics of science and to justify citation analysis.

3. The object–representation relationship between scientific literature and the *SCI* is assumed to result in an objective relationship between the reality of science and the results of citation analysis. This entails a translation of the notion of objectivity from it being engaged in the dualism objective—subjective to it being part of the polarity true—false.

In summary, the claim of citation analysis to objectivity and truthfulness is built on the *SCI* being different from, as well as identical to scientific literature. It is identical in as much as it uses elements of scientific literature and is consequently contingent on the patterns among these elements. Were scientists not referring to others regularly, a citation index would make no sense at all. This relationship of identity between the citation index and scientific literature is responsible for the index's objectivity: aren't scientists simply getting back what they have created themselves in the first place? At the same time, the index entails a drastic reduction of the complexity of scientific literature. In this difference between the index and its object lies the novelty of its representation of science. Moreover, the *SCI* gives the outsider's perspective on science. This external positioning contributes to its objectivity as well as to its novelty[11].

1.5 The quest for a citation theory

Because of the first two assumptions of citation analysis, the references of scientific articles — and only these — are supposed to be the building blocks of the citation index. Therefore, the citing behaviour of the authors of scientific texts has a direct relationship to the value of citation analysis. The latter must be accounted for in terms of the former, since the value of the citation is "ultimately grounded" (Chubin & Hackett 1990) in the referencing behaviour of the scientist. This has been the main paradigm from which the sociology of science has tried to construct a citation theory. One of the first systematic expositions of citing behaviour is provided by Robert Merton's sociology of science. It explains references in terms of the norms of science. Because of the constraining function of these norms, citing behaviour of scientists will display certain regularities:

[11] This is a general feature of the notion of scientific objectivity (Ashmore 1989).

> science is public not private knowledge. Only by publishing their work can scientists make their contribution (...) and only when it thus becomes part of the public domain of science can they truly lay claim to it as theirs. For that claim resides only in the recognition of the source of the contribution by peers. (...) The anomalous character of intellectual property in science (...) links up with the correlative moral as well as cognitive requirement for scientists to acknowledge their having made use of it. Citations and references thus operate within a jointly cognitive and moral framework. (Merton 1977)

Since the emergence of constructivism in the sociology of science, the act of citing has been analyzed in rather different ways. Empirical research has, moreover, revealed a bewildering multitude of motivations, functions and causes of references in scientific communication. Sometimes referencing is interpreted as the giving of credit where credit is due, sometimes as ways of persuading the reader, in other cases as merely perfunctory. The role of the reference, both in the citing text and with respect to the cited text, turned out to be equally varied. Scientometricians have repeatedly deplored the resulting lack of a proper and satisfactory theory of citing (Cronin 1984, Cronin 1981, Cozzens 1981, Cozzens 1985, Cozzens 1989, Luukkonen 1990, Luukkonen 1997), or analyzed the deficiencies of existing ones (MacRoberts & MacRoberts 1989, MacRoberts & MacRoberts 1984). Hence the call for "a citation theory":

> Not enough is known about the 'citation behavior' of authors - why the author makes citations, why he makes his particular citations, and how they reflect or do not reflect his actual research and use of the literature. When more is learned about the actual norms and practices involved, we will be in a better position to know whether (and in what ways) it makes sense to use citation analysis in various application areas. (Smith 1981)

Often the problem is felt to be the private nature of the act of citing:

> Logically, the use of citations as a basis for value judgements should imply that there is a universally recognized convention among authors. However, this convention, in so far as one can be said to exist, displays a remarkable resistance to standardization. (Cronin 1984)

In other words, the second assumption of citation analysis is partially fulfilled: scientists cite one another often enough to make a citation index feasible. It does not, however, legitimate citation analysis in the strict sense of a theoretically consistent scientometrical explanation. Within the scientometric community the practice of citation analysis lacks consensus about its theoretical foundations:

> we still have a theoretically underdeveloped understanding of what these bibliometric data actually mean. The continuous call for a theory of citation in quantitative science studies is itself indicative of the urgency to explore more systematically the relations between the use of scientometric methods and qualitative approaches in STS. (Leydesdorff 1987)

1.6 The reference and the citation

The attempts or rhetorical devices[12] to ground citation theories in referencing behaviour are based on the supposition that the reference and the citation are actually identical signs. But this is a tacit assumption, mostly hidden from view. When discussed explicitly, scientometricians seem perfectly aware of the difference between a reference and a citation. Price (1970) was the first to call attention to the distinction between the two signifiers, Narin (1976, 3) and Egghe & Rousseau (1990) later pointed to the same. The difference between the reference and the citation is, however, interpreted as a technical difference, hardly relevant for anyone but the inherently meticulous:

> If one wishes to be precise, one should distinguish between the notions 'reference' and 'citation'. If paper R contains a bibliographic note using and describing paper C, then R contains a reference to C and C has a citation from R (Price 1970). Stated otherwise, a reference is the acknowledgement that one document gives to another, while a citation is the acknowledgement that one document receives from another. So, 'reference' is a backward-looking concept while 'citation' is a forward-looking one. Although most authors are not so precise in their usage of both terms, we agree with Price (1970) that using the words 'citation' and 'reference' interchangeably is a deplorable waste of a good technical term. (Egghe & Rousseau 1990)

When authors expand on the distinction between reference and citation, they focus on the different characteristics of the distributions of references and citations. For example, Gilbert & Woolgar (1974) point to this (see also Chubin & Moitra 1975, Krauze et al. 1977):

> In a growing field, the characteristics (such as the average age and number) of the references in a paper will not necessarily be the same as those of the citations to a paper. The work of some studies is confused by giving both citations and references the same name. (Gilbert & Woolgar 1974)

Since these distributions of references and citations are not the topic of most scientometricians, the distinction Egghe & Rousseau (1990) refer to is glossed over most of the time.

A publishing author positions his text in a host of networks: a field-specific semantic one, a network of journals, an institutional network and so forth. The extraction of a citing network from the literature is, it should be stressed, one of the many possible representations of this literature. Whatever meanings the references have — we have already seen these can be very different — they are a striking feature of science. Their presence may even decide on the fate of the knowledge claims involved: "you can transform a fact into fiction or a fiction into fact just by adding or subtracting references" (Latour 1987). References share an important quality. Each reference is an inscription (Latour & Woolgar 1986),

[12]This depends on one's interpretation of the calls for a general citation theory: they can both be read as sincere attempts to construct such a theory, or as rhetorical moves to allay criticisms of citation analysis. These two ways of reading are not necessarily contradictory.

describing a certain text by a standardized code consisting of combinations of the title, author name, journal or publisher, year of publication, and page numbers. In other words, a reference, itself a piece of text, points to another text, the cited one. This does not mean, of course, that the latter can be found at the place the reference suggests. Since the reference is only a representation of the cited text, and not the cited text itself, the latter does not even have to exist. The reference can also be seen as a representation of what the author has read. Again, this does not have to be the case.

Irrespective of the way it is interpreted by the various theories of citing behaviour, the reference belongs to the citing text. Thus, the reference is completely defined by the citing text it belongs to and the cited text to which it points. In semiotic terms, the reference is a *sign* — a sign may be defined as the elementary unit of a representational system with the cited text as its referent.

The basic function of the *Science Citation Index* (and similar devices) is to turn an enormous number of lists of references upside down. Instead of organizing these references according to the articles they *belong* to, they are organized according to the articles they *point* to. If reference R of citing article A points to article B, the corresponding citation C is initially nothing else than a different format of reference R. The citation is the mirror image of the reference. This seemingly rather innocent inversion has important consequences. By creating a different format of the lists of references — by organizing the references not according to the texts they belong to, but according to the texts they point to — they become attributes of the cited instead of the original, citing, texts. Semiotically, the citing text is the referent of the citation. Hence the reference differs from its corresponding citation: the latter is produced from the former by inverting it. This inversion process is the basic symbolic act of producing a citation index and, actually, its fundamental operation. Without the inversion as the semiosis (creation of a new sign) of the citation, using the references to make an index would merely produce a reprint of bibliographies. The index would not be different enough from the scientific literature (its referent) to add information and thus be useful as a search instrument. The same inversion operation defines citation analysis. The basic act of citation analysis is a straightforward one: counting the number of times a text is referred to. Every citation analysis is based on counting the number of citations. The moment one starts to count citations of a cited text, one assumes this tells us something (whatever it may be) about the cited text or its position. Otherwise, the counting itself would be utterly pointless.

Thus, the giving of reference is one operation. The making of citation is a second one, reflexive towards the former as well as contingent on it. The shift in attribution of the two signs from the citing to the cited context is the crucial step. To be precise, it is also possible to stop the inversion halfway by attributing the reference to communication between the citing and the cited authors. In this case one would create a symmetric sign. One can then redefine communication as the process that is indicated by the reference[13].

[13]This is especially important if, as is often the case, the analyst does not have access to data concerning possible physical acts of communication between researchers. Narin's Influence Methodology (Narin 1976) is an example of this interpretation.

The inversion of a reference into a citation is a symbolic operation. It does not have to be embodied in any specific way; and if it does, this embodiment takes various forms. A citation index is not the inevitable result. For example, whenever a scientist "counts citations" by checking the bibliographies of articles looking for references to his work, he inverts references to citations. This private act of making citations does not amount to much, however. Only the production of publicly available citations counts. This is the reason that *being cited* is not exactly the same as *receiving a citation*. Although in principle every reference can be inverted to a citation, in practice many references do not enjoy this privilege. Citation indexes are therefore crucial in bibliometrics: only through these bibliographic devices do citations become publicly available, countable and therefore socially relevant.

It can therefore not be taken for granted that the scientist's desk would be the citation's birthplace by definition. A citation theory explaining citation analysis and related processes clearly cannot coincide with a satisfactory theory of citing behaviour. The act of citing is still important, but not because it gives the citation. Scientists are not so much creating the end product but more the raw materials of the indexing process. If a researcher or a scholar refers to a scientific article, he does not "give a citation", however odd this conclusion may seem, even to critics of citation analysis (Edge 1979). I do not mean to say that people using the term citation in its natural language meaning are wrong. In English, *citation* can mean both the sign citation and the sign reference. I am only saying that, if one wishes to analyze the citation, one should not be confounded by the beautiful ambiguity of natural language. It is essential to define the terms precisely.

The results of research into citing behaviour of scientists may still be relevant but cannot, contrary to received wisdom in scientometrics and science studies, be regarded as sufficient to explain the role and function of the citation. For this, the symbolic process at work in citation indexing needs to be analyzed. As noted, it is not a hard and fast rule that citations are ISI-made. Anyone can invert references. The problem is in the necessary scale of the process. It only makes sense to produce citations in large quantities. This makes the whole endeavour a complicated one, especially if one wishes to create a bibliographic instrument in science as a whole[14]. So far, this has only been done by ISI in Philadelphia, which explains the central position its databases occupy in the field of scientometrics.

1.7 The citation representation of science

To sum up, the citation is a new sign, different from the reference it builds upon. Its evolution and impact can best be understood if this feature is taken as the point of departure. Moreover, an individual citation does not have much impact. It is the ensemble of indicators that has shaped a novel representational form of science and technology, based on scientific literature.

[14]An inventor's perspective on the production of the citation index can be found in Garfield (1979, chapter 3).

Chapter 5 will analyze how these signs of science have been constructed[15] and have created a new type of representation of science and technology based on scientific literature. As has already been said, scientific literature is not a mirror of the research process. One can read *Science* or *Nature* to acquire knowledge about natural or social phenomena. They and more specialized journals can also be used to gauge the current situation in, for example, the sociology of violence or solid state physics. Researchers commonly read in both ways at the same time. The scientific literature is able to perform this double function because by representing the natural world (in the natural sciences), the social world (in the social sciences) or another world (in the humanities and engineering sciences), it immediately also represents scientific research.

To put it in more general terms, the literature is a heavily stylized representation of science and technology, focusing on cognitive claims and recontextualizing these (Knorr-Cetina 1981). Science's substantive claims and results are the core around which the literature is organized. The scientometric indicators mentioned above are rooted in literature. They capture various relationships between publications. But, and this is a crucial point, they ignore their content. The scientometric representation of scientific literature is built on its formal properties. The ensemble of indicators generate a re-representation of science and technology which expressly ignores the cognitive dimensions involved. As a result, two very different representations have come into existence. This study explores the interactions between these two representational forms of scientific and technological research in science policy and science studies.

The act of representing is not just an intellectual but always also a political intervention (Haraway 1991, Hacking 1983, Rouse 1987, Hagendijk 1996, Sassower 1995). Scientometricians play at politics by creating a specific image of the sciences they analyze. This study is made political by representing scientometrics as being one of the principal embodiments of the new citation culture. The question is not whether the scientometric representation is a "correct" or a "false" one, but how it differs from the representational format based on science's substantive claims and results. This difference creates new options for science policy. Chapter 6 traces the early introduction of the scientometric representation via science and technology indicators in Dutch science policy. This chapter tries to sketch the emergence of a policy market for science and technology indicators by looking in detail at the introduction of these indicators into science policy in the Netherlands. Science policy is commonly analyzed in terms of conflicting and converging interests. I hope to show in the concluding chapter (chapter 8) that science policy can perhaps more fruitfully be analyzed as an intricate interplay between various science representations. Using this perspective, the citation culture can also be located more precisely: at the interface between science and politics. The citation culture is a hybrid; it is both political and scientific. Citation indicators and scientometrics may deal mostly with rather dry technical, instrumental and methodological issues, nevertheless they have created an irreversible transformation in the politics of science.

[15]This chapter is contrary to the preceding two not historical in character but is a conceptual deconstruction of the most important indicators.

Although the profession of scientometrician is a highly political one, science policy was not the first domain to encounter the new sign citation. Sociologists of science, like Robert Merton and Derek Price, recognized the potential of the *SCI* long before most science policy officials had even heard of it. Chapter 4 will analyze the interaction between the emerging citation culture, the "science of science", and the sociology of science. It is certainly not meant to be an attempt to write the history of the science of science. The chapter merely looks in somewhat more detail than present histories of the sociology of science have done, at the way the *SCI* has affected the emerging science studies. The new index seemed to be significant for sociology in two respects. First, it was a rich source of new data for the sociological study of science. Second, it promised to make an old dream come true: the application of "the scientific method" to science itself, an idea central to the science of science. This approach entailed the idea of rationally analyzing and managing scientific research to increase the interaction between science and society (Bernal 1939). The invention of the citation index with its "objective" data seemed to enable a more objective analysis of science than was previously feasible. The scientific method could finally be applied to science itself! Since science was conceived as a more objective form of knowledge than history for example, the seductiveness of this prospect to people like Derek Price can hardly be overstated. They embarked enthusiastically on the *SCI*. This led to the merging of the older tradition of the science of science with the citation culture, the result being scientometrics. In other words, chapter 4 tells the story of the birth of scientometrics as a distinctive scientific specialty.

Scientometrics appears to be a hybrid specialty of social science. It is located at the interface of science proper and science policy. It produces indicators as policy instruments. In other words, scientometrics is one of the regulatory sciences (Jasanoff 1990). What these specialties have in common is that their development is strongly shaped by the regulatory process. Moreover, scientometrics is aimed at regulating science itself, at least partly. It is therefore a reflexive regulatory science. Building upon the analysis in chapter 4 and chapter 6, chapter 7 tries to profit from this reflexive capacity by applying scientometrics to itself in order to understand the socio-cognitive evolution of scientometrics.

1.8 Representing scientometrics

This study hopes to provide an informative account of the citation culture. I realize that this analysis differs from the usual perspectives on citations entertained by scientometricians, science policy officials, sociologists of science, or researchers in general. Scientometricians feel they are measuring science, either as "scientists of science" or as sociologists. For science policy people, scientometrics is just one of many sources of policy instruments. Scholars in science studies tend to view scientometrics merely as a method without theory. Lastly, scientists tend to be divided into two groups: opponents and supporters. This is also true of researchers in the social sciences and the humanities. Adversaries raise all sorts of arguments against measuring science in general (e.g. the unmeasurable creative nature of scientific discovery) and citation analysis in particular (e.g. the lack of meaning of the citation). The proponents of citation analysis tend to see the scientometric scrutiny of the scientific process as a means of improving the quality of research, notwithstanding its limitations.

This account of the citation culture cuts across these divides. I present this representation as a stand-in for the usual image of scientometrics. This study will try to show that the citation culture as a hypothetical construct can explain the dynamics of measuring science more elegantly than either of the aforementioned more usual accounts. This does not invalidate the stories of these actors. I hope that scientometricians and science students will recognize part of themselves. At the same time, all competing analyses of scientometrics mentioned earlier are somehow actors' accounts[16]. This may give the impression that my account is somehow more encompassing than theirs. After all, their stories are part of mine whereas my story is not included in theirs. I am the analyst, they are the actors. My arguments may therefore seem more powerful. In short, I could present my analysis as a birds-eye view of scientometrics whereas the actors are inevitably myopic.

Unfortunately, I cannot in all honesty present my case this way. As the author of this study I cannot escape also being an actor in the field I am studying. By analyzing scientometrics I belong to the specialty of science studies. This is just one of many studies in the field of science studies. I cannot claim any special privilege for my analysis compared with those of my colleagues. This is also true of my colleagues' accounts of scientometrics. But these are the very competitors mentioned earlier! Thus, my analysis does not have any privileged position vis-à-vis theirs. This is not all, however. Scientometrics is also part of science studies. Scientometrics, like science studies in general, is capable of analyzing science. It can also, again like science studies, analyze itself. Such an analysis is included in chapter 7. This study is therefore both scientometric and non-scientometric. Thus, I am not only an actor in science studies in general, but also in scientometrics in particular. To be more precise: I am an actor precisely *because* I am an analyst. Apparently, the usual distinction between actor and analyst breaks down. This is the very issue around which the discussion about reflexivity revolves in science studies, an issue that will again be touched upon in chapter 8.

As has already been said, this representation of scientometrics is inevitably also a political intervention. I do not side with or fight against the citation culture. This study is not a polemic but an analysis. This does not mean, however, that I would be a neutral intermediary or a reconciliator. I do not strive for consensus in these matters. The most I can hope for is that this study will give a new perspective which can be used as a resource by whoever wishes so. Science studies are politically relevant since they deconstruct a powerful source of knowledge and a source of power. How to deal with this political dimension in our analyses remains a matter of contention. All possible positions on the spectrum between a strict separation of science and politics (Collins 1991) and explicitly siding with one party (Scott, Richards & Martin 1990) have been defended. Hagendijk (1996) has pointed to the reflexive nature of this debate: the fact that methodological

[16]This may often be the case in contemporary history.

neutrality has political consequences (and is therefore not neutral) does not mean that siding with one party in a controversy is less problematic. The analysis of the citation culture you are about to read, deals with these issues by incorporating the controversy over measuring science for policy (mostly by citation) in the analysis itself. This entails a translation of political questions into analytical ones, which is itself a political act (Sassower 1995). As a consequence, old political choices may be seen in a new light. Whether — and if so how — the new analytical questions should in their turn be translated into political matters is up to you as the reader of this study.

Back to Contents Page

Back to Contents Page

Chapter 2

The creation of the Science Citation Index

2.1 Mixed reception

NEWS RELEASE
Contact: Mrs. Joan E. Shook
INSTITUTE FOR SCIENTIFIC INFORMATION 33 SOUTH SEVENTEEN STREET PHILADELPHIA 3, PA.
phone/locust 4-4400 cable/currcon twx/ph 803

For Immediate Release

$300,000 GRANT TO PROBE INFORMATION RETRIEVAL AWARDED TO INSTITUTE FOR SCIENTIFIC INFORMATION BY NATIONAL INSTITUTES OF HEALTH AND NATIONAL SCIENCE FOUNDATION...

THREE YEAR PROJECT TACKLES CITATION INDEX TECHNIQUES FOR SCIENCE

Research scientists will soon be consulting a more precise and specific literature index that links together subject material that would never be collated by usual indexing systems. Concerned with new starting points for scientific literature searches, the unique concept uncovers sometime-buried associations, relating important works and authors, yet keeps the researcher abreast of the masses of current published scientific information. This new approach to information retrieval is called the Citation Index.

A $300,000 grant extending over a three-year period has been awarded to the Institute for Scientific Information, Philadelphia, Pennsylvania, to study the practicability of citation indexes and to test their techniques of preparation. The project, under joint sponsorship of the National Institutes of Health and the National Science Foundation, is aimed at producing a unified citation index for science including the publication of a genetics index.

Two years after this press release[1], the *Genetics Citation Index* was published (Garfield & Sher 1963). It was quickly followed by the first volume of the *Science Citation Index* proper (Garfield 1963). Since then, the *SCI* has been published four times a year. Nobel Prize winner and Stanford University geneticist Joshua Lederberg[2] wrote the preface to the Genetics Citation Index. He emphasized, as the press release in 1961 had done, the potential of the *SCI* as a tool for the scientist: "Citation indexing can uncover unexpected correlation of scientific work that no other method could hope to find, and a succesful match can often be located with great speed and assurance"[3]. Yet, the new tool was not generally applauded. "Citation indexes have had a mixed reception", judged library scientist John Martyn on 31 March 1965 at the evening meeting of Aslib in London. According to his estimation, based on personal contacts, scientists were on the whole in favour of citation indexes, while librarians were much more cautious (Martyn 1965, 188). The latter tended to compare the citation index with the bibliographic tools they were used to, and did not find the new method quite so advantageous. As one library scientist noted somewhat regretfully: "Citation indexing (...) has been imposed upon us" (Shank 1965). According to Martyn this divergence can be explained by the differing information need: "The librarian is concerned with information *retrieval*, whereas the scientist is more interested in information *access*; regarded as a retrieval tool, the Citation Index is not as efficient as some more conventional approaches to the literature, but as an *access* tool it functions very well" (Martyn 1965, 189).

A number of reviews in scientific journals were however positive. "This *Genetics Citation Index* represents a landmark in literature-searching in genetics" wrote the *Eugenetics Quarterly* (Anonymous 1964*b*). A colleague agreed:

> At $ 100, the price may seem high—but to those who can use this index, it could pay for itself in time alone, to say nothing of its value in supplying new leads. This latter will probably be its chief value, for, unlike other indexes which supply only the information specifically sought (if that), citation indexing can lead in unforeseen directions and to unforeseen relationships. (Anonymous 1964*a*)

Geneticist J. A. Beardmore saw the *SCI* as "a real effort to help those people endeavouring to follow specific lines of thought and endeavour through the rapidly increasing bulk of articles concerned with genetics and related fields" (Beardmore 1964)[4]. Not every scientist shared this appreciative outlook, though.

[1]For the complete text see the appendix.

[2]Lederberg received the Nobel Prize in 1958 for his work on the bacterial genome.

[3]Lederberg also praised the inventor of the *SCI*, Eugene Garfield: "My own contribution to the project has been too limited to inhibit me from commending Dr. Garfield and his associates for organizing and implementing a project which has required an unimaginable attention to detail, technical skill, enthusiasm, and above all, an irrepressible concern for meeting the real need of scientists. To flourish, science has many needs but none are more vital than responsible communication with history, society, and posterity embodied in what we casually call the scientific literature." (Garfield & Sher 1963, iii) This chapter will show that Lederberg's role in the citation indexing project was anything but limited.

[4]The enormous amount of work involved in the building of the citation index was commonly praised. Heinisch spoke of "bewundernswertem Fleiß" (Heinisch 1965).

Nature's reviewer was more critical of the *SCI*:

> As it is, for the physicist, this index covers only about 5 per cent of the 800 journals included in the *Physics Abstracts* for 1961. While it would be unwise to underestimate the possible value of this method of indexing, I cannot visualize many situations where these volumes could be used more effectively than other indexes. (Cleverdon 1964)

One year later, the author judged the cumulative *SCI* for the year 1964 (Garfield 1965) "an improvement in many ways" (Cleverdon 1965). But he still found the future potentialities of the *SCI* "difficult to assess". In 1966, *Nature* was even more critical:

> Have the bad old days when it was only possible to get on by knowing the right people gone for good? Alas, no. An instrument recently introduced requires just this kind of intellectual nepotism, yet without the necessary personal acquaintance. (Anonymous 1966)[5]

Some readers could not agree less. A "regular user" of the *SCI* reacted:

> I suspect that you have not used the *Index* for its proper purpose. (...) S.C.I. permits the questioner to exploit fully his own intelligence and experience in his analysis of the literature without having to rely on the arbitrary choice and use of key words by abstracters. (Davies 1966)

Physicist John Ziman attacked the lack of the physics journal *Philosophical Transactions*, the extremely small print of the index and the fact that the first *SCI* did not give the titles of the citing articles. The first regular issue in 1964 was improved in several of these respects (including the citing titles) and ISI promised a rapid increase in journal coverage.

Whereas *Nature*'s Cleverdon reasoned partly along the lines of the librarian, *Science* had the *SCI* reviewed by a zoologist, in an experimental way. Steinbach[6] reviewed the *SCI* "as a novel and interesting example of a device to cope with the scientific literature explosion" (Steinbach 1964). He came to a moderately positive judgement[7]. The *Genetics Citation Index* did not yet include titles, it was mainly a list of author's names (cited and citing authors). A trial run which he conducted with some graduate students resulted "in a split vote" on the question "whether the virtue of completeness compensates, to the scientist, for the work necessary to relate authors' names to scientific content" (Steinbach 1964). The overall judgement was positive: "Reading about and trying to use the *Index* has convinced me that it will be of value and that all scientists should indeed examine it and consider its potentialities" (Steinbach 1964, 142). Nevertheless, the author

[5]This publication was a short note about the first quarterly volumes of the *SCI* for 1966.

[6]The reviewer was professor of zoology and chairman of the department of zoology at the University of Chicago; he served as chairman of the Division of Biology and Agriculture of the National Academy of Sciences-National Research Council from 1958 to 1962 (Steinbach 1964).

[7]"Any real evaluation of *Science Citation Index* must be based on an extensive use test, and there has not been time for that" (Steinbach 1964).

"could not agree completely" with Garfield's statement that both the usefulness and the desirability of citation indexes was already proven (Garfield 1964): "This statement may be correct with respect to its use by administrative personnel and librarians; its usefulness in advancing the wisdom of science must be judged in the future by scientists"[8].

The difference of opinion between Steinbach and Garfield relates directly to the role attributed to scientific literature. Steinbach was critical of Garfield's enthusiasm for collecting all information in one system. "There is something wistfully comforting in the thought of knowing all facts. It is an old idea that, if the scientific facts are known, wisdom follows" (Steinbach 1964). He did not share this idea. In his opinion, the more basic the research and the more important the idea, the less important the literature. Hence, he found it "hard to think that Newton, Lavoisier, or Loewi" would have used a citation index (Steinbach 1964). The main point of contention was, however, the broad range of possible uses of the *SCI* advertized by ISI (Institute for Scientific Information 1964). These concerned, amongst others, the evaluation of the impact of research, sociological research, and studies into the literature use by scientists. Steinbach warned against these uses:

> Misused, some of these "other applications" could cause some important difficulties, tending to foster the idea that what has been good in the past is best for the future. (A "high impact factor" means more support?) (Steinbach 1964)

Two years later, *Science* expressed a much more favourable opinion, notably on the policy aspects in its lead article by Philip Abelson, the journal's editor:

> A particularly useful tool in a search for significant articles in a particular field is the Citation Index. (...) An interesting by-product of the Citation Index is a new method of evaluating scientific productivity. Instead of counting a man's reprints, one counts citations of his work by others. Already sociologists are examining the value of this new analytical tool. They note some limitations but find that a citation index is a valuable aid to management. (Abelson 1966)

In summary, the record of initial reactions to the first publications of the *SCI* confirms Martyn's (1965) conclusion that librarians were sceptical (Martyn 1965, Martyn 1966). Scientists seem, however, to have been more divided than Martyn acknowledged at the time. Some were quickly sold on the new possibilities of roaming the literature of their own accord. Scientists who focused less on keeping abreast with the literature were probably indifferent[9]. "Not *every* scientist greets the appearance of a Citation Index with enthusiasm", Martyn also noted in 1965 (Martyn 1965). Scientists as well as librarians who gave the *SCI* some consideration, seem to have been especially wary of the possible use science policy.

[8]Note that Steinbach, in constrast with Martyn, thought that librarians would be more appreciative than scientists!

[9]Joshua Lederberg, Interview, February 3, 1992, New York.

This was explicitly mentioned in the leaflets and brochures of the Institute for Scientific Information, the publisher of *SCI* (Institute for Scientific Information 1964), as well as in Garfield's publications (Garfield & Sher 1966) from the very beginning[10].

These mixed feelings were, nevertheless, far more positive than the reactions Eugene Garfield had experienced in the preceding years. He had been actively propagating the idea of a citation index for science since he had become with it in 1953. Few had responded. Even a few years before getting the grant from NIH and NSF, referees of his proposal were quite critical, some even hostile[11] (Anonymous 1959 or 1960):

- It is my firm conviction that what is needed most urgently at this time is not more components research (e.g., citation index) but a prototype system of scientific communication that can serve as a test bed for existing and proposed components and that will guide their development. .... To the extent that Mr. Garfield's proposal does not promote the growth of an experimental test environment, to that extent I consider it unwise and wasteful.
- As a necessary step, presumably, the applicant will use non-scientists to scan the literature. This step can be very harmful. A citation index must be overcautious about errors.
- I fail to see how this sort of publication (i.e., through journals) even as an 'intermediate mechanism' could be of much value to geneticists.
- The problem of avoiding a flood of citations to routine, expected, ordinary, usual, humdrum references, which nobody in the world would ever want to consult, is one that should be solved before a grant is made favorable to this proposal.[12]

This resistance to the idea of investing in citation indexes of scientific literature may come as a surprise to the present-day user of the *SCI* and *SSCI*. After all, scientists must acknowledge their peers and must share their ideas and resources with their colleagues (chapter 1). Therefore, it seems rather obvious to use the footnotes of a scientific article, or the bibliography of a book, as an entry in a literature searching procedure. The use of citation frequencies, the number of times an article is quoted, may seem rather straightforward too. The citation score of an article may be seen as a measure of its use by other researchers and therefore of its impact, importance or quality. Not coincidentally, Eugene Garfield considers citation scores to be the condensed peer review of the entire scientific community (Garfield 1979).

[10] I will come back to the relevant policy debates later in this chapter as well as in chapter 6.

[11] An undated and anonymous overview of the referees comments (Anonymous 1959 or 1960) can be found in Garfield's Personal Archive, Phildelphia. I have not been able to locate these documents in NSF's Historian's Archive, most documents having been destroyed. November 9, 1964 Garfield sent this document to his collaborator Irving Sher with the following note: "Did you ever see this? Must have come to me when I first applied to NSF for support five years ago." To which Sher replied: "Oh gosh! The problems you had to overcome!!"

[12] Some comments were positive, while most dealed with technical details (Anonymous 1959 or 1960).

The fact of the matter is, however, that the concept of the citation index did not come to science as naturally as this interpretation would suggest[13]. Without Garfield's perseverence and his strong belief in the usefulness of a citation index, we would probably live in a world without ISI's bulky volumes and, of late, shining disks. The creation of the *SCI* is less the outcome of some inevitable process in science than historically contingent. To start with, the *SCI* has its roots not in science but in law.

2.2 Enthusiasm for citation

2.2.1 Shepard's

Citation indexes were already old hat for American lawyers at the time this history of *SCI* starts. In the second half of the nineteenth century, one Frank Shepard in Illinois deemed it useful to know whether a legal proceeding was still valid. He produced gummed paper with lists of cases which cited the case in hand. Lawyers in Illinois glued them into their dossiers so enthusiastically that in 1873 Shepard set up a commercial business. His company, Shepard's Citations Inc.[14], had the monopoly on producing the one and only citation index "To serve the Bench and Bar". First in Chicago, later in New York and then in the 1950s in Colorado Springs, a staff of highly qualified lawyers produced the Shepard's Citator by hand, covering all judicial decisions in the United States. Shepard's was a respectable firm, proud of its supreme reliability[15]. Its product was grounded in the norms and procedures of the legal system. As W. C. Adair, former vice-president of the company, explained to the readers of *American Documentation* in 1955:

> The lawyer briefing a case must cite authorities to back up his arguments. So must the court in writing its opinions. This is because of the doctrine of "Stare Decisis" which means that all courts must follow precedents laid down by higher courts and each court generally also follows its own precedents. (...) The lawyer, however, must make sure that his authorities are still good law, that is, that the case has not been overruled, reversed, limited or distinguished in some way that makes it no longer useful as a valid authority. Here is where the use of Shepard's Citations comes in. (Adair 1955)

The searching procedure was simple. First, the lawyer located a case similar to his own, then looked up Shepard's citator to see whether later cases had cited it. He would immediately see whether the precedent was still valid[16] and which other cases had made use of it. Adair told his audience that important law suits were won "on the strength of a case located by the use of Shepard which no other method of research disclosed" (Adair 1955). In summary, "Shepardizing" legal literature has since 1873 been based on the authority-centred norms of the United

[13] As will be shown in this study, the *SCI* has profoundly changed its meaning and function throughout its history.

[14] Shepard's was taken over by Reed Elsevier and The Times Mirror Company on July 3, 1996.

[15] "To Serve You Better" was the slogan of an advertisement of Shepard's Citation Index in 1954.

[16] A small r in front of the case meant for example that it was reversed.

States' legal system: the most recent decision of the highest court is valid[17]. The way of indexing by citation perfectly tied in with this value system. One can hardly think of a sharper contrast with supposedly ruthless scientific criticism. This hierarchical indexing style served nevertheless as the model for ISI's *Science Citation Index*.

2.2.2 Adair

Retired from Shepard's, running a cattle ranch in Colorado Springs, but still eager to work, William Adair read in the local newspaper sometime in 1953 an article stating that the scientific world "was being swamped in a sea of literature". It was a report on *The First Symposium on Machine Methods in Scientific Documentation* organized by the Welch Medical Indexing Project at John Hopkins University in Baltimore. This project had been sponsored by the Army Medical Library since 1948 (Miller 1961, Larkey 1949). The main task of the project was to find out whether machines could be used to improve the efficiency of indexing and retrieving medical literature, and if so how. The indexing itself was supposed to be the tried and trusted subject indexing. In this respect the Welch Medical Library was not very innovative. Within these boundaries, the staff had to devise new systems of indexing, subject-heading and ways of using machines to solve "the literature problem" and was organizing meetings in the country. Adair decided to write a letter to the supervisor of the project, Sanford Larkey. He told Larkey about the citation indexing system, informing him of his opinion that "if the whole body of American Law can be classified so that a knowledge of one case can be used as a key to locate all other cases in point, the same thing can be done with medical articles"[18]. Adair offered his expertise: "I have retired from Shepard's and am now free to undertake and organize such a project". He got a reply of a twentyfive-year old junior member of the staff, named Eugene Garfield. Garfield did not know anything about citation indexing. He wrote Adair that his suggestion would be investigated, but kept him at a distance. "We do not have any positions open for staff members", Adair was told[19]. Nothing happened. Adair's initiative had no impact on the Medical Indexing Project.

Only more than a year later, after he had been fired by Larkey[20] [21], did Garfield resume contact with Adair "with the idea of writing a paper to be published in one of the learned society journals"[22]. Having browsed through *Shepard's Citations* at the public library, the idea had begun to appeal to him. Garfield had even written a paper on "Shepardizing the scientific literature" for his professor while he was a fellow at Columbia University (Garfield 1954*b*)[23]. At first, Garfield was

[17]This is true for every juridical system in which all citizens are equal before the law.

[18]Adair to Larkey, March 10, 1953.

[19]Garfield to Adair, March 16, 1953.

[20]Eugene Garfield, Interviews, Philadelphia, January 27, 1992 and February 4, 1992.

[21]In the first half of 1954, Garfield joined the literature section of the pharmaceutical company Smith, Kline & French as "consultant in machine documentation" (Garfield to Adair, June 11, 1954; Garfield to Adair, October 7, 1954; Garfield to Adair, August 24, 1954).

[22]Garfield to Adair, June 11, 1954.

[23]Garfield wrote Garfield (1954*b*) for professor Fleming while he was the Grolier Society Fellow

not sure if citation indexes could be indeed applied to science:

> Without knowing exactly what you had in mind I do not feel it is fair for me to be discouraging at the outset. But the one thing that must be kept in mind when comparing the field of science with that of law, is that there are anywhere from one to three million articles each year appearing in the scientific journals.[24]

At that time, Garfield was not yet thinking of building a citation index. Working as a consultant in automation, he focused on possible uses of computers[25]. He perceived a possible opportunity in automating the production of citation indexes. Adair "was very glad" to receive Garfield's letter[26]. He did not think the "vastness of the field" affected the problem at hand "except insofar as it bears on the expense and labor required". Whether a citation index of medical literature would be effective, depended in his view on "how much a certain article cites other authorities": "In the legal field it is rare that one finds a case on statute no matter how old that has never been cited"[27]. A citation index would provide "a chain which would automatically string together authorities alike probably not all". Adair was less certain about potential investors:

> There is no lawbook publisher of any size, and I know them all, who would be in a position to make the enormous investment necessary and would have the extra plant and equipment. To form a new company to undertake the

at Columbia University Library School. This was probably sometime before April 1954, since the draft of his 1955 article for *Science*, which was based on his Library School paper, is marked "Submitted in April 1954" (Garfield 1954*a*).

[24] Garfield to Adair, June 11, 1954.

[25] "My interest in writing you is merely as an individual who is concerned with the difficulty of managing our scientific research record. My own specialty is the utilization of machines in facilitating the compilation of compendia such as Shepards. It is not unlikely that the Shepard Company already has mechanized many of its procedures, but my first reaction would be that an enterprise such as the Shepard citation system would find many uses for punched-card machines or large scale computers, commonly referred to as "electronic brains", an appelation which has probably been detrimental in their quick utilization by such companies as Shepards. (Garfield to Adair, June 11, 1954). With the term "electronic brains", Garfield probably refers to Shaw (1949), in which the librarian of the US Department of Agriculture Ralph Shaw discusses the "rapid selector". This was an electronic device using a photocell which enabled rapid selection and printing of documents stored on microfilm. Originally a German invention, it was further developed by Vannevar Bush at the Massachusetts Institute of Technology. Bush and later Shaw devoted much attention to this device because it promised to improve "the quality of organization of knowledge both for administrative routines and for communication among scientists" (Shaw 1949), apart from the reduction in storage space needed. Shaw (1949, 169) also discusses whether this microfilm-reader was a thinking machine: "The selector has been termed a 'thinking' machine or 'electronic brain'. Without more knowledge of what 'thinking' consists of, it is difficult to say whether the selector thinks or does not think. Certainly in the common sense of the term the selector is not a thinking machine. It merely stores vast amounts of data and sorts and reproduces them in accordance with instructions given to the machine both in the coding and in the selection. All the machine ever does is to match black dots (or, if you prefer, light dots) with complementary dots in an interrogating card."

[26] Adair to Garfield, June 21, 1954.

[27] Adair to Garfield, June 21, 1954.

> risk involved would not be feasible. To my mind it could be done only in one way. That would be to interest one of the large foundations on the ground of increase in the dissemination of scientific knowledge.[28]

He was willing to write an article but was not sure whether it would be understood by scientists[29]. After some encouragement[30], Adair sent Garfield a rough draft[31] which Garfield forwarded to Jesse Shera, editor of *American Documentation*, who readily accepted the paper[32]. Only now did Garfield admit that he had submitted a paper (Garfield 1954*b*) to *Science* in April[33]. Adair was enthusiastic about Garfields text: "I have read your article and think it's ideal to follow mine. (...) I was quite surprised at your grasp of how the work might be done"[34]. Adair's article appeared in the January 1955 issue of *American Documentation*[35].

[28] Adair to Garfield, June 21, 1954.

[29] "I could write an article and would be glad to but since my experience has been in the legal field, I'm afraid an article on the citation phase of legal research wouldn't be too understandable to those outside the legal profession." (Adair to Garfield, June 21, 1954.)

[30] In his reply, Garfield proposed they write an article together: "In any case, my original thought was that possibly you and I could join forces to compose an interesting article that would be comprehensible to scientists, de-emphasize the legal citation phase, except by way of reference and illustration" (Garfield to Adair, June 26, 1954). Again, Adair responded positively:"I do have time to write an article on the citation phase of research and would be glad to do so if you think it would be worthwhile. I agree that the legal phase should be emphasized only for purposes of illustration and to suggest how citations might be applied in other fields. I would think that you might add whatever you thought might be appropriate from the index angle. At any rate I should be glad to have you make whatever changes you thought might improve the article since you are closer to the field of scientific research than I. If you approve, let me know and I'll get busy on it at once" (Adair to Garfield, July 22, 1954). Garfield encouraged him and ensured him of its usefulness: "An article on the citation phase of research could be a great contribution. As an associate editor of the journal *American Documentation*, I am sure it would be published. I would be glad to read it over or help in writing it. Perhaps the best way to start is for you to write a first draft (let me handle the typing, i.e. my secretary) and on the basis of this I can get a better idea of what you have in mind and let you know if we have the same basic problem in mind" (Garfield to Adair, August 24, 1954), and promised to send him a copy of his earlier paper: "Professor Fleming at Columbia University has still not returned to me the paper I did for him on "Shepardizing the Scientific Literature", but I am writing him again so that I can send it on to you." (Garfield to Adair, August 24, 1954)

[31] Adair to Garfield, September 18, 1954.

[32] "I have heard from Dr. Shera and as I expected he is delighted to have your paper for publication in American Documentation". (Garfield to Adair, October 7, 1954)

[33] "Now that you have submitted your paper I am sending you a paper I wrote while I was on a fellowship at Columbia University as I wrote you in my letter of June 11, 1954. A while ago I sent this paper to a colleague at Johns Hopkins University, Professor Bentley Glass, who is a member of the editorial board of SCIENCE the publication of the American Association for the Advancement of Science (AAAS). Should it be accepted for publication it is possible that some foundation may see the value of a citation system in disseminating scientific knowledge—as you mentioned in your letter of June 21, 1954. Ther (sic) is small chance that my article could be published before next year sometime. That is why I feel your article will make an excellent start, especially since American Documentation is widely read in scientific circles. When my article is revised I will then be able to make reference to your paper. Any suggestions you may have that will improve my paper will be greatly appreciated—just indicate the comments on the manuscript in your wonderful green ink." (Garfield to Adair, October 7, 1954)

[34] Adair to Garfield, October 11, 1954.

[35] Its title was proposed by Garfield. Originally, Adair had titled it "A Citation System for

The author stressed the potential of the citation index as a novel way to unlock the body of scientific literature, a point which would be raised again and again by Eugene Garfield in the years to come:

> The amazing efficiency of the citation method is such that once the starting case or statute is found it becomes a key that unlocks the entire store of law on a given point. It is this function which it appears would be of great value in other fields. An article on any scientific subject would be the key to all others. It may be objected that a comprehensive index would do the same thing. Even then the vast number of titles, sub-titles, cross references, etc. make the most skilfully compiled index difficult to use for the purpose of exhausting a subject. (...) The index represents the opinion of the compiler a to where a given subject should be pigeonholed. The list of citations is essentially determined by the authors, i.e., the courts. (Adair 1954)

He also anticipated objections "that whereas legal cases and statutes have standard references, scientific articles do not", and the problems stemming from the "enormous amount of scientific literature". Adair closely followed the practice in law in dealing with these potential problems. The references could be standardized[36] and the vastness of the task could be made smaller by "splitting up the field of science broad by subject matter such as chemistry or medicine and by restricting the number of years covered". The main point, according to Adair, was that of the citing culture in science: "More important than these is the question as to how much do writers on scientific subject cite other writers and articles. The writer must assume that they do this to a considerable extent"[37]. (Adair 1954).

2.2.3 Computers

By being made privy to Adair's personal experience with citation indexing, Garfield became more enthusiastic "with each passing day"[38]. The use of computers was the crucial perspective Garfield added to Adair's way of thinking about citation indexing. They discussed the possible applications of punched card systems and computers in considerable detail. Adair did not think much of it, though. He told Garfield that it would not be very easy to use computers:

Scientific Literature". Garfield did not think this would be clear "since the expression "citation system" is also used to mean the method for expressing bibliographical citations" (Garfield to Adair, October 18, 1954).

[36]"Shepard's covers many law reviews and journals and some special publications such as the Journal of the American Patent Society. These are given abbreviations and an abbreviation table is shown in the front of the books." (Adair 1954)

[37]Adair refers here to the introduction of his article: "In order to clarify the scope and purpose of this article it is perhaps well to explain that the writer was for many years the executive Vice President of the Frank Shepard Company, publishers of Shepard's Citations, a system of legal research used with great success by lawyers and jurists for over three quarters of a century. In the course of his incumbency he has seen occasional requests from members of the medical and engineering professions for information and advice as to whether such a system might not be used in their special fields. Unfortunately no one connected with the Shepard Company had the time to go into these questions thoroughly".

[38]Garfield to Adair, October 18, 1954.

> Concerning your idea of the possibility of interesting the Shepard people in mechanical processes I might say that we went into this very thoroughly back in New York. The Hollerith people put their best man on the job and I worked with him for over a month. We finally had to give up for two reasons. First the cards didn't have enough spaces to cover the vast bibliography involved. Second and more important each citation would require a card and since citation slips are handled by the millions the punching machines and trained operators required was prohibitive. In addition to this very detailed attempt we twice had very good industrial engineering firms make surveys and their results also were mil [nil] on the production side.

This did not deter Garfield:

> To answer your comments concerning your previous experience with the Hollerith (IBM) people — I am sure that the problem here is by no means a simple one, but on the other hand my own experience with IBM has time and again taught me that they do not have the best talent available for every type of problem, and indeed are often beat at their own game so to speak.[39]

Adair insisted. In his reply he informed Garfield about the way Shepard's produced its citator. An army of highly competent lawyers and clerks checked and double-checked the validity of every citation. They did everything by hand. It was necessary for the slips on which the citations were written to be legible. Adair was of the opinion that punched cards, incomprehensible by nature, would introduce too many errors into the process[40]. Garfield immediately sent him a sample of a medical index page prepared with IBM cards, proving that legible information could be printed on them[41]. He asked Adair whether the latter would not recommend him to Shepard's, and explained how he would proceed in advising them as a documentation consultant. Adair did not, however, wish to be explicitly involved: "If you feel like writing to the Company, do so. But please do not refer to me. Mr. W. G. Packard who is still the President knows that I worked on this problem for years and that I found it unsolvable. He would wonder a great deal if I now recommended a further survey"[42]. Adair was con-

[39]Garfield to Adair, October 27, 1954.

[40]Adair to Garfield, October 30, 1954.

[41]Garfield to Adair, November 12. 1954.

[42]Adair to Garfield, October 30, 1954. From the very beginning, Adair had not wished to attract Shepard's attention to his discussions with Garfield. "I attended the company picnic recently and I frequently play golf with some of the younger executives and editors but I haven't mentioned our correspondence to any of them as yet. Later on it may be advisable to do so. (...) It would only lead to a lot of correspondence and inquiry at a time when our ideas aren't sufficiently crystallized." (Adair to Garfield, October 11, 1954.) Later, he even asked Garfield to destroy his letters: "I would appreciate it if you would make whatever notes you care to from this letter then destroy it. You might do this with previous letters too. This may seem foolish to you, but Shepard's enjoys an absolute monopoly in its own chosen field and its' interior workings are guarded jealously. I'd really like to recommend you to the company but I don't see how I could do it without revealing the idea that brought us together. Since I'm supposed to be retired the use of the company for other citation purposes would seriously damage my present friendly status." (Adair to Garfield, "Tuesday". Between November 12 and 24, because on that day Garfield replies to this letter. (Garfield to Adair, November 24, 1954.))

vinced by the punched cards Garfield had sent him, but insisted that automation was still highly questionable. His main argument was the editorial work needed[43]. Garfield nevertheless sent the Shepard firm his draft article for *Science* (Garfield 1954*a*), as well as a series of detailed questions about the way the citator was produced[44]. L. A. de Bow, executive vice-president of Shepard replied politely without answering all questions[45]. He objected to Garfield using the word "Shepardizing" in his draft article[46], and informed Garfield that "the basic idea of a citator for scientific literature is, as you of course know, not a new one": "We have ourselves often considered such a publication and have, on numerous occasions over the past 25 years, discussed with doctors the possibility of a citator covering medical literature". By then, Garfield was already in possession of this information. As early as 1947, Shepard's had been approached by an officer of the American Medical Association requesting the company to undertake a citation system covering medical literature[47]. But the Shepard management did not feel it had the resources to produce any system other than in its own legal field. Somewhat later the company produced a sample system for the American Society of Electronic Engineers. However, as Adair later recalled, this society did not have the money to go on with it[48]. In 1954, as before, Shepard's did not venture beyond its juridical domain, Garfield's light prodding notwithstanding[49].

[43]"Would the machines save enough time in the assorting, which is small in the current work, to show a reduction in cost on this one operation? This is the point on which I have doubts. The big cost factor in the Shepard work is the editorial work which must be done by lawyers. (...) there is, of course, no mechanical means for doing this mental work. This cost has been the impasse for all engineers the company has employed" (Adair to Garfield, "Tuesday" See note 42.)

[44]Adair did not expect much from this: "If you feel like making a proposal to them I suggest that you do so by all means. The last survey was by a firm of young industrial engineers who thought they could apply the same methods they would use in a steel plant. It was a fiasco which cost the company a large amount with no return." (Adair to Garfield, December 2, 1954.)

[45]"the time required to do so would be much greater than we can devote to such a tabulation" (De Bow to Garfield, January 4, 1955).

[46]"We are certain you will understand our position that "Shepardizing" is a trade term relating to the use of Shepard's Citations only" (De Bow to Garfield, January 4, 1955)

[47]Adair to Larkey, March 10, 1953. Adair to Garfield, March 24, 1953.

[48]Adair to Larkey, March 10, 1953; Adair to Garfield, June 21, 1954.

[49]To Garfield's question of whether Shepard's had lost interest in citators outside of the field of legal literature (Garfield to De Bow, January 17, 1955), the answer was no: "The fact that we have, as I told you, several times in prior years considered citators to scientific literature, particularly to publications in the medical field, does not mean that we are no longer interested in extending our activities to those fields. On the contrary we have that possibility very much in mind but to date have been sufficiently occupied in completing citators for all the juridical jurisdictions and simply have not come to any conclusion on any other subject" (De Bow to Garfield, January 24, 1955). De Bow was curious about Garfield's punched cards but in the end concluded that "As of this time, we do not feel that punched-cards alone offer any advantage over the methods we use" (De Bow to Garfield, May 18, 1955). They exchanged a couple of more letters, but nothing came of it.

2.2.4 Patents

In the mean time, Garfield had been looking for possible applications of a citation index[50]. His first idea was to look into the potential of a citation index for patents[51]. His idea was to submit a formal proposal to the Patent Office. He asked Adair to be one of his consultants. Adair was very interested in this[52]. He also informed Garfield that, again, he was not the first to think of a patent citation index[53]. Patent attorney Harry Hart of Bell Telephone Laboratories had spoken with Adair in 1945 about "Shepardizing the Proceedings of the Institute of Radio Engineers"[54]. The firm had at the time been "too busy moving out of New York" to do much about it[55]. In 1947, Hart suggested US Patent Office Commissioner Ooms that he adopt the same system as the lawyers (Hart 1949)[56]. Two years later, patent lawyer Arthur Seidel of Gulf Oil Company published the same idea in the *Journal of the Patent Office Society* (Seidel 1949), explaining that a citation index "would not require much effort upon the part of the Patent Office" whereas it would accelerate the search of comparable patents. Hart immediately took up the issue once again, for the first time hinting at the notion of a citation network: "It furnishes a network of paths which cut across the major highways

[50]"After giving a great deal of thought to the possible use of Citation Indexes as outlined in my paper, I pondered where such a system would be particularly useful. To make a long story short, I quickly realized that a wonderfully fertile application is in the area of patents. I immediately checked to see the type of patent coverage Shepard's already has and found that it concerns only the legal phase, albeit a very important one. (...) Since patents form a very substantial and important part of the scientific literature, I could not help but be impressed by the possibility of Shepardizing the patent literature. I immediately did a sample of the work necessary and find that the practices of patent examiners make the application of a Citation Index an extremely pertinent one because there appears at the end of every printed patent specification a list of patents "cited" as well as any other pertinent literature references. These cited patents are provided by the patent examiner and not the patentee, making them extremely useful and pertinent to any future patent search (Garfield to Adair, October 18, 1954).

[51]Garfield made an appointment with Mr. Lanham of the U.S. Patent Office, with whom he was "extremely well acquainted".

[52]"In regard to your using my name as a consultant on anything pertaining to citations, feel free to do so. I could devote almost unlimited time right now. (...) Your technical knowledge on scientific literature and indexing is far beyond mine. We ought to supplement each other nicely. (Adair to Garfield, "Thursday". This letter is undated but is written on October 21, 1954, as is clear from Garfield to Adair, October 27, 1954.) .

[53]"As regards your idea for patents, I know that patent lawyers would welcome it. We used to get many requests from patent firms for such citations. At the time we revised the Federal Reporter Citations I spent considerable time both by correspondence and conferences with patent firms and officials of the patent office. I found that such citations would relieve patent lawyers of the necessity of a trip to Washington to examine the file wrapper or of obtaining the information through a correspondent Washintong firm. Here is the reason we didn't go into it. If we put such a section in the Federal it would bulk the book tremendously and the majority of subscribers not being patent lawyers would feel they were paying for something they didn't need. Why not a separate publication? We took a list of all the patent lawyers in the U.S. From this we estimated the number of firms, then added the big law libraries. The total subscribers we could hope to get wasn't sufficient to cover the cost" (Adair to Garfield, "Thursday" See note 52.).

[54]Garfield to Adair, October 27, 1954.

[55]Adair to Garfield, October 30, 1954.

[56]The suggestion was forwarded to the Shepard firm.

marked by the (inevitably artificial) boundaries of Classification" (Hart 1949)[57]. Yet, nothing came of it. Garfield was not discouraged by these failed precedents. Together with the Atlantic Research Company he submitted a proposal to the Patent Office[58], after having attended a meeting of the Advisory Committee on Patent Office Mechanization[59]. By May 1955, however, nothing had happened yet. The Patent Office had set up a new Research and Survey group which, Garfield thought, "is considering the proposal that was submitted"[60].

2.3 The citation introduced to science

2.3.1 *Science*

The energetic Garfield, who was by now firmly sold on the concept of citation indexing, hoped that the articles Adair and he were going to publish would attract attention. The reactions to the publications in 1955 in *Science* and *American Documentation* were, however, disappointing. True, Garfield received several letters endorsing the idea and encouring him to develop it further[61]. Generally, however, resistance seemed to prevail. *Science* published two letters from experts in scientific documentation, both unfavourable to Garfield's plea. "Amid today's overwhelming difficulties in scientific communication (...) this index would solve too few problems to justify its surely great cost at this time" (Schoenbach 1956)[62]. This counter plea for more inclusive subject indexing was strongly supported by a second letter (Welt 1956)[63]. In *Science*, Garfield's proposal was not discussed any further for the next few years.

This is remarkable, because in his article Garfield tried to tie in his proposal, as directly as possible, to the running debate about scientific documentation. He opened by referring to earlier articles on the "uncritical citation of critized data" (Thomassen & Stanley 1955, Zirkle 1954)[64]:

> In this paper I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete

[57] Hart was quite clearly enthusiastic about citation indexing: "It permits a later student to reap the fruit of an earlier student's effort. It helps to redirect the stream of classification into the most suitable channels. In short it is a splendid and unique system, serving in a way that no Index or Classification Manual can possibly serve." (Hart 1949)

[58] Garfield to Adair, November 24, 1954.

[59] Garfield to Adair, November 12, 1954.

[60] Garfield to Adair, 4 May 1955.

[61] Garfield to Allen, February 15, 1957.

[62] The author was with the Literature Research Division of the National Drug Company, Philadelphia.

[63] The author was with the Cardiovascular Literature Project, Chemical-Biological Coordination Center, National Academy of Sciences-National Research Council, Washington D.C.

[64] The accuracy of scientific data was at the time a priority in the documentation discussion. This opening section was not yet included in the draft paper (Garfield 1954*a*). The second important difference between the draft of October 1954 and the published text in July 1955 is the disappearance of the term "Shepardizing". Garfield deleted this term from the title and the text at the request of Mr. De Bow of Shepard's (See note 46..) The other changes are mostly minor corrections or reorganizations of the text.

> data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers. It is too much to expect a research worker to spend an inordinate amount of time searching for the bibliographic descendants of antecedent papers. It would not be excessive to demand that the thorough scholar check all papers that have cited or criticized such paper, if they could be located quickly. The citation index makes this check practicable. (Garfield 1955, 108)

The index would be very handy for the working scientist: "It is best described as an association-of-ideas index, and it gives the reader as much leeway as he requires". In this respect, the citation index would be, Garfield stressed, far superior to the traditional subject indexes which by nature restrict the interpretation of the article to a predefined number of topics:

> One of the basic difficulties is to build subject indexes that can anticipate the infinite number of possible approaches the scientist may require. (...) What seems needed, then, in addition to better and more comprehensive indexes, alphabetical and classified, are new types of bibliographic tools that can help span the gap between the subject approach of those who create documents—that is, the authors—and the subject approach of the scientist who seeks information. (Garfield 1955)

Not only did Garfield focus on the information needs of the scientist, he also translated the concept of the citation index in terms of the subject indexes with which both scientists and librarians were more familiar.

> Over the years changes in terminology take place, that vitiate the usefulness of a standard subject index. To a certain extent, this is overcome through the citation approach, for the author who has made reference to a paper 40 or 50 years old has interpreted the terminology for us. By using author's references in compiling the citation index, we are in reality utilizing an army of indexers, for every time an author makes a reference he is in effect indexing that work from his point of view. This is especially true of review articles where each statement, with the following reference, resembles an index entry, superimposed upon which is the function of critical appraisal and interpretation.

His experiences at the Welch Medical Indexing Project, as well as his exchanges with Adair were utilized in the discussion of the coding of entries and the way the citation index would look[65]. Although he hinted at the possibility of giving the index the appearance of a bibliography[66], Garfield thought of the index as an ordered array of numbers. Each article would be represented by a two-part code (the first part referring to the journal, the second to the article). Under each cited article, the citing articles would be printed, with a one-letter classification

[65]These experiences are also reflected in the mentioning of the review articles about which a great deal of discussion had taken place at the Welch Medical Indexing Project.

[66]"Thus, it would be possible to list all pertinent references under each case with sufficient completeness to give the index more of the appearance of a bibliography. However, this would result in an extremely bulky volume." (Garfield 1955, 108–109)

Citation Index Entry	
11123s-687	
	464-9789(R)
	869-3366(R)
	1105-9876(A)
	1123-4432(R)
	a11,123-0752(O)
	-0779(O)
	-7264(O)
	-7331(O)
	-7385(O)
	-0866(O)
	-8221(O)
	-9158(O)
	-9497(O)
	-9529(O)

Figure 2.1: The citation index example Garfield presented in *Science*.

added to indicate the nature of the citing article ("an original contribution, review article, abstract, and so forth") (figure 2.1 on page 32). Garfield described the production process with punched cards, stressing that "relatively unskilled persons can perform the necessary coding and filing". The citation index would amount to a complete listing of all articles that had referred to a specific publication. He mentioned several possible uses of the index, amongst which "an individual clipping service" that the citation index could easily provide and "the tracking down" of the origins of an idea. The first application Garfield mentioned was, however, "historical":

> This would clearly be particularly useful in historical research, when one is trying to evaluate the significance of a particular work and its impact on the literature and the thinking of the period. Such an "impact factor" may be much more indicative than an absolute count of the number of a scientist's publications (Garfield 1955, 109)[67]

2.3.2 Propaganda

Garfield, now an independent documentation consultant[68], was not deterred by the silence that followed his proposal. He undertook several initiatives to make the citation index more popular, which increased his grip on the intellectual and practical difficulties in compiling this type of index. The editor of *Chemical Bulletin* invited him to write a short article (Garfield 1956*a*). This article stressed the citation index's foundation in experience[69]. It is moreover the first time Garfield

[67] Garfield refers here to citation studies (Gross & Gross 1927, Brodman 1944, Fussler 1949) that had been inspired by the problems librarians had in dealing with the growing literature.

[68] Garfield advised amongst other the pharmaceutical company Smith, Kline & French.

[69] "The beauty of the citation index is that it achieves *a posteriori* indexing because a citation is *experiential*. The "logic" of all conventional scientific classifications has inevitably broken down with experience. Artistotelian logic has been a chain around the neck of the scientist and classifier

directly linked his ideas about the citation index to the plea for a centralized clearing house by the famous crystallographer John Desmond Bernal in 1948:

> Bernal proposed some time ago that a centralized reprint clearing house be established. Each scientist would then regularly receive papers in designated areas of interest. The proposal is excellent in its simplicity. Its execution is not so simple. How would one spell out his interests? (...) However, a reprint distribution plan based on the principle of the Citation Index could overcome this difficulty. (...) It would, in effect, be an individual clipping service.

Together with Mrs Margaret Courain, supervisor of the Research Files Division at Merck, Garfield produced an experimental citation index of patents, which he presented at the Mineapolis meeting of the American Chemical Society on 16 September 1955 (Garfield 1957)[70]. He informed the gathering that he had submitted a feasibility study to the Patent Office:

> Some time ago a formal proposal was submitted to the Patent Office, suggesting that a feasibility study be conducted. If anyone is interested they can question the Patent Office on the matter. Pending action by the Patent Office, which seems doubtful, I am hopeful that Chemical Abstracts[71] and/or Shepard's Citations may be able to undertake the compilation of a Citation Index to patents.

Apparently, although Garfield was by now systematically campaigning for a citation index in different areas, he still did not wish to produce it himself[72]. His main business was the publication and development of *Current Contents*, which was a success from the outset, and the exploration of innovative information products for scientists, medical doctors and engineers. For example, January 1956, he submitted a "tentative proposal" (Garfield 1956*c*) to the National Science

alike. Since the Citation Index is an arbitrary construct rather than a "logical" one, it can stand the test of time. Citations are permanent and unique, as are the works they identify. The significance of men's writings may change, but their identities are fixed." (Garfield 1957)

[70] In this presentation, Garfield again stressed the point of view of the literature users: "Every thought, idea or discovery can take on new meaning depending on the user's frame of reference. (...) The Citation Index breaks this "barrier" by presenting subject matter in Bibliographical arrays which are neither alphabetical nor classified but associative". (Garfield 1957) This was especially relevant to patents, since "Citations are provided for a variety of reasons, but principally to disqualify certain claims. Citations are also made frequently in order to restrict the scope of the applications." Garfield pointed to the wastefulness of not using a citation index: "The lack of action in the Patent Office is particularly unfortunate since a system of this type for the exclusive use by the examiners could include references to abandoned applications which are not classified at all and which contain a wealth of citation information insofar as the citations resulted in abandonment. It is not unlikely that many searches have been repeated by examiners in part or in whole when an examination of the appropriate abandoned file would turn up much useful data. Examinators do often remember these abandoned files. However, it is difficult to see why they have to rely on their memories for abandoned files and not for published patents. Classification is used precisely because the human memory has its limits".

[71] Chemical Abstracts had announced that they were studying the potential of a citation index.

[72] Sometime after the meeting, Garfield was told that although the Patent Office considered the citation index "a basically good idea", it concluded that "the time required to compile and use it would not justify acceptance of the proposal" (Garfield 1957).

Foundation for a grant of the order of $ 50,000 a year to determine the information requirements of scientists[73]. In May, Garfield submitted a proposal for research on mechanical indexing[74]. NSF's reply was characteristic of its general attitude towards private firms in scientific documentation: "When we discussed it on the phone, I tried to make it clear that the chances of approval would be better if it were submitted by an institution, in this case, the University of Pennsylvania"[75]. The grant agencies NSF and NIH would in the future repeatedly make the same point with regard to Garfield's citation indexing proposals. Garfield returned the favour by being very critical of NSF's policy and its way of handling scientific documentation needs. At the December 1955 meeting of the American Association for the Advancement of Science in Atlanta, he made a strong plea for a centralized national documentation centre (Garfield & Hayne 1955)[76]:

> The National Science Foundation does perform a few functions of such a center, but it is essentially a government counterpart of other philantropic foundations that encourage and support scientific research. NSF has no centralized documentation apparatus. (Garfield & Hayne 1955)

Partly as a personal exercise, Garfield prepared a citation index of the Old Testament which he presented in 1956 to the American Documentation Institute in Philadelphia[77] (Garfield 1956*b*). In this talk, Garfield presented a new idea, interpretative citation indexing. He sketched two possible approaches available to the indexer: "One he can limit himself to collecting all formal citations indicated in the text. (...) A second approach would be interpretative citation indexing. Based on the subject matter disclosed in the text the indexer himself can provide citations that relate passages to what has been published elsewhere. (...) This type of interpretative indexing is to be compared with exegesis." (Garfield 1956*b*). The similarity with the procedure of a patent examinator is not coincidental[78].

2.3.3 Allen

January 1957, Garfield received the first serious support from a scientist. "Dear Mr. Garfield", geneticist Gordon Allen[79] wrote,

---

[73]Garfield to Brownson, January 25, 1956; Garfield to Brownson, October 29, 1963.

[74]Brownson to Garfield, July 18, 1956.

[75]Helen Browson, author of this letter, was NSF's Program Director for Scientific Documentation. (Brownson to Garfield, July 18, 1956).

[76]In their presentation, Garfield and Hayne linked documentation to the prevailing fascination with spying: "documentation is the forerunner of intelligence". Said Garfield: "The CIA's scope is totality of information pertinent to the nation's safety and progress. The time has come for an analogous scientific body." (Anonymous 1956)

[77]Garfield to Packard, February 22, 1956.

[78]Garfield also stressed in this presentation the importance of very precise references: "Thus, in a citation index there would be given the specific page of the citing as well as the page or passage of the article cited. This is extremely important, because the citation index is designed to deal with complex thoughts usually not revealed by the seemingly specific subject headings assigned by indexers".

[79]Allen was at the time with the Department of Health, Education, and Welfare of the National Institutes of Health.

> Since the appearance of your article in Science two years ago, I have been eagerly looking for some news of steps toward a citation index. I have urged the American Society of Human Genetics to take some initiative in the matter[80], but they are already involved in the construction of a subject index in human genetics.
> The references I have seen to your suggestion (for want of a citation index, I probably have not seen all of them) have been disappointingly cool, and I wonder if you have received any personal letters that were more enthusiastic.
> If a group of interested persons were brought together, they might be able to make some headway.[81]

Garfield's reaction shows the impasse he had reached with regard to the citation index:

> It was most kind of you to write me concerning your interest in citation indexes. I wish that I could report that we were well on the way towards preparing a comprehensive citation index to scientific literature. This is far from the case. However, I am more than ever convinced that a citation index is an absolute necessity in this era of voluminous publication. In my own personal reading and literature searching, I constantly see the need for citation indexes.[82]

Garfield sent Allen a copy of the paper he had published in the *Chemical Bulletin*[83]. He also told Allen that he had had "frustrating" experiences with the National Science Foundation[84]. More than a year later, Garfield heard that NSF was planning to support research on citation indexes[85]. He promptly inquired who was going to do the research and was told[86] that he could submit a proposal[87].

[80] Allen had done this in 1956, when he talked to Sheldon Reed, then president of the society. (Allen to Macklin, April 9, 1959)

[81] Allen to Garfield, January 24, 1957.

[82] Garfield to Allen, February 15, 1957.

[83] "I couldn't have picked a better publication for burying an idea, but the former editor was kind enough to ask me to write the article which is more than I can say for some of the more "learned" journals." (Garfield to Allen, February 15, 1957).

[84] "I had planned to submit some sort of proposal to the National Science Foundation for a grant that would enable me to continue research on citation indexes. However, I have found it very time-consuming and frustrating to deal with NSF and could not afford the time needed to prepare lengthy proposals. However, I would be delighted to get together with you or any other interested persons to work out ways and means of preparing a citation index to the literature of genetics and/or any other discipline." (Garfield to Allen, February 15, 1957).

[85] This statement seems to have been the result of a congressional hearing in which a congressman asked NSF why it didn't shepardize the literature. As a result, NSF published the statement that it was considering the investigation of "a proposed method of bringing related material together similar in some respects to Shepard's Citations, a respected method in the field of law, which has never been tried in the sciences" (*Hearings, House of Representatives, Committee on Appropriations* 1957). One of Garfield's friends drew his attention to this statement (Garfield to Lederberg, May 21, 1959).

[86] M. M. Berry to Garfield, August 22, 1957. This communication, to which Garfield referred in his research grant application of July 15, 1960 (Garfield 1960*a*), was probably a phone call.

[87] Garfield recalled this episode a year later in his first letter to Joshua Lederberg (Garfield to Lederberg, May 21, 1959) and summarized it in his research grant application of July 15, 1960

Together with Gwen Bedford[88], a researcher at the University of Pennsylvania, he prepared a draft feasibility study of citation indexes (Garfield 1958*a*) in May 1958 and sent it to NSF for commentary[89]. At the same time, he wrote a paper for the upcoming International Conference on Scientific Information and informed Allen of both actions[90]. Allen reacted sympathetically: "Your Unified Index to Science sounds like a tremendous advance over present indexing, and I hope you can get it started"[91]. He could not, however, invest much time himself:

> I am still interested in a citation index for science, but it does not appear that I can give any of my own time to it. The American Society of Human Genetics had already embarked on its own subject index when I brought up this matter, and is still too engrossed, I think, to consider any broader approach. However, I will tell them, at the next annual meeting, about your present plans and invite their interest and support. I'm afraid you can't look to them for direct financial help, but is there something else they can do, as a society or as individuals, to help push the project along?

Two weeks later, Garfield officially submitted his proposal to NSF (Garfield 1958*b*). Its goal was "to determine the utility of citation indexes for science in terms of general usefullness, invariance in time, minimizing the citation of poor data, identification of the "impact factor", and provision for individual clipping services". The study should also develop "a suitable technical design for citation indexes". Its motivation followed Garfield's (1955) line of reasoning, albeit with an added emphasis on the index's potential for "the encyclopedic integration of scientific statements"[92]. Garfield moreover re-emphasized the potential of Bernal's proposal for a central clearing house[93]. This document also makes

(Garfield 1960*a*).

[88]Gwen Bedford was with the Institute of Cooperative Research of the University of Pennsylvania. She had been associate professor at the School of Library Research, Drexel Institute. It was thought that she could possibly use the results of the project for her PhD dissertation, which would concern "the impact of the government research report on conventional scientific communication" (Garfield 1960*b*, 5).

[89]Garfield sent it on June 25 and heard on July 7 that Program Director Helen Brownson found it "suitable for submission to the Foundation". She also expressed as her opinion that the salary levels "seemed quite high to us, for the Foundation ordinarily supports research and studies at universities where salaries are considerably less" and asked Garfield "to trim the budget as much as you can". (Brownson to Garfield, July 7, 1958).

[90]Garfield sent Allen his paper and told him that he had "worked up" a research proposal (Garfield to Allen, July 24, 1958).

[91]Allen to Garfield, July 30, 1958.

[92]"The citation index for science can be the key to the "encyclopedic integration of scientific statements", the basic synthesis which can weld disciplines and specializations. Operations research studies have established beyond dispute the value and considerable power of the results of the cross-fertilization of disciplines in the search for solutions to complex problems. We are now at the point where we need a tool which is suitable for tackling scientific problems on a broad front. The citation index is such a tool." (Garfield 1958*b*, 3)

[93]"From time to time, the idea of establishing a centralized reprint clearing house has been advanced (cf Bernal 1948). Each scientist would regularly receive reprints in designated areas of interest. The citation index offers an ideal basis for determining areas of interest which are realistic and practical. As a scientist's frame-of-reference changes, so does his distribution list — and without lengthy exchanges of correspondence." (Garfield 1958*b*, 4)

clear that Garfield's "impact factor" was meant to be a tool for cultural historians mainly, and should not be read as an antecedent of present scientometric impact factors:

> The index provides for a direct qualitative (and possibly a semi-quantitative) measure of the "impact factor", i.e., and evaluation of the significance of a particular work and its impact on the literature and thinking of the period. This is the tool cultural historians need so that their studies can be rigorous in method and thus lead to meaningful results. (Garfield 1958*b*, 4)

The project was meant to be a two-year study, starting September 1958, which would be restricted to compiling the index. Garfield asked Allen to phone the office of Scientific Information of NSF[94] in support of his proposal[95]. Allen phoned Burton Adkinson at NSF's office of Scientific Information, then explained his reasons in more detail in a follow-up letter:

> The greatest value of a citation index, I think, would be its ability to connect any scientific statement with subsequent amendments, corrections, or retractions.[96] (...) Another great service that a citation index would perform would be to make readily available the latest work on any specialized topic. (...) In summary, a citation index would, I think, do for our present system of science indexes what cross-references do for any simple index. By adding another dimension, it would increase tremendously the efficiency of any literature search.[97]

One month later, NSF turned down Garfield's proposal, but at the same time expressed its interest in a citation index[98]:

> In view of the interest in citation indexes as such expressed by the reviewers we believe that the actual preparation and subsequent testing of one might be desirable. It seems to us, however, that the scientific subject field of the experiment should be carefully defined either by a group of scientists or by a scientific literature group closely associated with a specific subject area. We think the preparation and testing of the index also should be under the cognizance of such a group. Therefore, I would like to inquire what your reaction would be to a possible situation in which a scientific society or literature group in a technical field appropriate for citation indexing would provide scientific direction for the experiment, and your organization, with its interest

[94]"Some of the people in the National Science Foundation need this kind of prodding in evaluating projects and I am sure that a favorable opinion from a reputable scientist like yourself would certainly do us no harm when our proposal comes up for consideration." (Garfield to Allen, August 15, 1958)

[95]Garfield had created his own company by now, Eugene Garfield Associates with offices in Spring Garden Street in Philadelphia.

[96]Allen gave some examples of "self-perpetuating errors" in the human genetics literature.

[97]Allen to Adkinson, September 5, 1958.

[98]"Dear Gene: By now you will have received the Foundation's letter saying that it cannot support your research proposal for a general feasibility study of citation indexes for science. The purpose of this letter is to ask your reaction to a possible modified approach to the problem." Gray to Garfield, October 23, 1958.

> and experience with citation indexing, would prepare the actual index. Such an approach, it seems to us, would combine the skill of the group most acknowledgeable in the techniques being studied with the scientific judgment and counsel of representatives of the "public" expected to use and benefit from the index.

2.3.4 A World Brain

Garfield took this as a flat refusal, proving once again NSF's inability to deal with the tasks at hand. He did not stop his campaign, though. In November, he made his plea for a "unified index to science" at the National Academy of Sciences conference on scientific information (Garfield 1959). In this presentation[99], the idea of integrating scientific knowledge, already mentioned in Garfield (1958*b*), was further developed:

> Since scientific research today is highly inter-disciplinary, the "selective" approaches of our traditional media, based on the old academic disciplines, can never give us anything more than makeshift tools, which do not function properly, considering the overall job to be done. Fragmentary approaches are not only inefficient but inadequate. (Garfield 1959, 674)

A unified and standardized approach to scientific literature searches was Garfield's goal:

1. Provision for *one* logical *starting point* for all literature searches, regardless of subject.
2. *Standardization of nomenclature*, particularly in the areas of overlap between existing indexing services.
3. Provision for *detailed indexing* not possible in specialty indexes. An increased number of analytical entries per article would be economically and intellectually more feasible.
4. Elimination of all doubt as to whether individual articles had been indexed by specialty indexes, particularly in inter-disciplinary subjects where selectivity exercised by specialty indexes is necessarily arbitrary. *Complete coverage of all individual articles* becomes a practical possibility.
5. *Economic utilization of machines* for the compilation of the present specialty indexes and indexes to individual journals.
6. *Economic production and distribution* of scientific indexes by virtue of broadening the number of potential users. Mass production is the best known method for reducing product costs.(Garfield 1959, 675)

[99]Garfield's paper grew out of an earlier paper he prepared while being at the Welch Medical Indexing Project, entitled "Unified International Scientific Indexes through Centralized Machine Indexing and Its Relations to Standardization of Nomenclature". (Garfield 1959)

For Garfield, such a unification of indexes would mean a step towards realizing H. G. Well's "World Brain"[100]. In fact, this idea of a large, global information system was not only Garfield's paradigm but also that of other major players in the history of the *SCI*. Garfield's 1958 presentation was the first in which he explicitly put his citation index idea in this wider perspective: "The Citation Index has been discussed in previously published articles. However, this is the first time its use for consolidating references to and from the various abstracting services has been recommended. This feature of the Unified Index is significant" (Garfield 1959, 676).

2.3.5 Lederberg

In April 1959, Gordon Allen again tried to convince his American Society of Human Genetics[101] to co-operate with Garfield to create a citation index. He told president Madge Macklin that Garfield's submission to NSF provided "a rich opportunity for Human Geneticists to get a valuable free service"[102]. Allen, who had been one of NSF's reviewers of the proposal, informed Macklin of the reasons Garfield's proposal had been turned down and urged the society "to appoint a panel of interested and competent members" to consult with Garfield. He also proposed recommending the proposal to NSF. One month later, Garfield received a letter from another World Brain enthusiast, geneticist Joshua Lederberg, which would prove the turning point in the history of *SCI*:

> Since you first published your scheme for a "citation index" in Science about 4 years ago, I have been thinking very seriously about it, and must admit I am completely sold. In the nature of my work I have to spend a fair amount of effort in reading the literature of collateral fields and it is infuriating how often I have been stumped in trying to update a topic, where your scheme would have been just the solution! I am sure your critics have simply not grasped the idea, & especially the point that the author must learn to cooperate by his own choice of citations + thus he does the critical work. Have you tried to set this out in an adequate experiment? Would you look for support from the NSF? Of course you have to count on opposition from the established outfits, which have already succeeded in blocking any progressive centralization of the Augean tasks.[103]

As he later explained to Garfield[104], Lederberg's initiative was prompted by a science policy debate in the Genetics Study Section of NIH. The administration wished to evaluate its actual impact on research, and proposed, in the words

[100] "A unified index to science could take many physical forms. In a large centralized science information center, this H. G. Wells type of "World Brain" might be a 3 by 5 inch card file, a random access electronic storage device, or a searching device such as Minicard or Filmorex. In this paper an alphabetic printed index is assumed." (Garfield 1959, 676)

[101] Allen was member of the board of this society.

[102] Allen to Macklin, April 9, 1959.

[103] Lederberg to Garfield, May 9, 1959.

[104] Lederberg to Garfield, July 29, 1960.

of Lederberg[105] "a number of rather fancy and inefficient schemes". Lederberg recognized that a citation index would accomplish the purpose "at a negligible additional cost" and decided to contact Garfield. The latter was very happy to receive this letter:

> I hope you won't be embarrassed by a show of emotion, but your memo almost brought tears to my eyes. It then seemed that over six years of trying to sell the idea of citation indexes had not been completely in vain.[106]

He told Lederberg the whole story of his pleas for citation indexes, the support of Gordon Allen and the resistance he had met since 1954:

> As to opposition from the established outfits—there is no end to this. Chemical Abstracts pays lip service to Citation Indexes, but does nothing about them. Even my friends at Biological Abstracts and the Current List of Medical Literature who accept my judgement on many other conventional problems—look upon Citation Indexes as something impractical and unnecessary—particularly when there is so much more abstracting and regular indexing left undone.[107]

He did not hesitate to inform Lederberg of his impressions of NSF, recalling the episode the year before:

> I tried to convince certain illiterates at the National Science Foundation to give me a small grant to conduct research on citation indexes. (...) Needless to say my proposal was turned down.

Lederberg was shocked by Garfield's letter and "absolutely astonished that citation indexes are not long since a standard feature at the Patent Office"[108]. He adviced Garfield to resubmit his proposal "to all the agencies who could be interested"[109]. Especially NIH "would be an excellent target", since it "is anxious to evaluate its 'impact' on scientific progress, and how better do this than through your scheme"[110]. Garfield did not agree with Lederberg's judgement of the rea-

[105] Lederberg to Garfield, July 29, 1960.

[106] Garfield to Lederberg, May 21, 1959.

[107] Garfield to Lederberg, May 21, 1959.

[108] Lederberg to Garfield, June 18, 1959.

[109] "I imagine your tactical error was in approaching the NSF at a time when, to put it bluntly, it was too broke to do more than a fraction of what it should. It is still not so well financed, and I think you might have better luck with another agency. In fact, the way to push it is to submit the same proposal, with due notice, to all the agencies who could be interested. These would include NSF, NIH, AEC, ONR, AFRDC.... (Lederberg to Garfield, June 18, 1959). Lederberg also thought that the propects would be better in biology or medicine than in chemistry; that Garfield should stress the job needed mainly money and machines, not professional manpower; and that starting with a review journal would not be such a bad idea.

[110] Lederberg was himself member of a NIH panel with this task. During one such meeting, he recalled Garfield's article in *Science* and decided to write the man (Joshua Lederberg, Interview, February 3, 1992, New York).

sons for NSF's lack of support[111]. Neither did he share Lederberg's still sympathetic attitude towards the Patent Office[112]. He even poured cold water on Lederberg's suggestion of contacting NIH: "I think that I anticipated you on the idea of getting the NIH Div. of Research Grants interested. Their former librarian, Scott Adams, tried to get them interested but nothing came out of it". Lederberg reacted promptly to the materials Garfield sent him along with his pessimistic assessment of potential support from the science funding agencies. What Garfield had taken as a flat refusal (section 2.3.4 (page 38)), he read far more positively: "I find myself rather more sympathetic with the viewpoint summarized in Dwight Gray's letter of 23 Oct 58, and which I would interpret as a constructive basis for further dealings on your part"[113]. Lederberg's interpretation was different from Garfield's because the geneticist concurred with the reviewers' doubts about the proposal. He wondered whether "in fact the concept hadn't already been well enough sold to the NSF reviewers":

> My own feeling at the present time is that the utility and feasibility of citation-indexing are, in fact, self-evident; it is rather doubtful that any limited sample would serve to convince anyone else who did not already see the point. (...) I can easily see that $ 59,000 might be thought a wasteful expenditure if its main effect were to reprove the obvious, and especially if not very much more than this would be needed to get a useful product.[114]

2.3.6 Re-establishing communication

Lederberg proposed Garfield "jump in" and ask for NSF's assistance in organizing a scientific committee as suggested by NSF's last letter on the subject. At the same time, he tried to convince Garfield that not only "dolts" but also quite recep-

[111] "I can't agree that in this instance the reason for the turn down was the financial condition of NSF. (...) in the Office of Scientific Information they go around pleading that nobody wants to do research in documentation and always have. What they mean is that nobody wants to do the kind of research they want. (...) They also give out money for "popular causes" like translation of Russian stuff—regardless of its scientific value. You can't imagine how frustrating it has been in the past five years (or maybe you can) to have had at the helm of scientific documentation activities in NSF a woman who was neither a scientist or an information specialist, but just a good secretary (a Spanish major) who worked her way up by taking good notes at meetings and preparing reports for her bosses. I would never say this publicly, but that is the absolute truth. I tried for five years to get some kind of support so I wouldn't have to go "commercial" but it was closing battle. I even got myself temporarily affiliated with the Univ. of Pa. ICR and the Franklin Inst. and couldn't make a dent. (Garfield to Lederberg, June 23, 1959).

[112] "I am sorry but your are trying to give the Patent Office people credit for more intelligence than they have. You don't know how backward they are. It is such a tradition bound organization that even their approach to machines, which they are investigating, is completely archaic. I suggest you meet their Dir. of Research some day if you want to be convinced. They did not reject the Citation Index on the grounds you suggested—it was purely on the grounds that they didn't think it was worth the effort. (...) And Congress wonders why it takes over two years to get patents and sometimes longer.

[113] Lederberg to Garfield, June 26, 1959.

[114] Lederberg to Garfield, June 26, 1959.

tive people worked in the funding agencies[115]. In other words, Lederberg tried to teach Garfield how to handle these institutions. The latter assured Lederberg that he was not crying wolf: "I think you will find, if you haven't already, that I am really a very reasonable man and that I do not have a persecution complex about the NSF"[116]. He accepted Lederberg's critical judgement of his proposal[117], was "heartily in favor of forming what you call a "consumers group", and acted accordingly by re-establishing contact with NSF's Program Director for Publications & Information Services Dwight Gray:

> Dr. Lederberg suggests that I ask your assistance in establishing a group of scientists who would advise us on just exactly what course should be pursued in regard to the development of citation indexes. Since there has been interest indicated for the genetics literature, I would like to suggest that this field be the "test" field and that possibly another field could be tested simultaneously. This second field might be fishery biology as the FAO, Rome had indicated a definite interest[118] in helping out in this.[119]

He also informed Gray that the American Society for Human Genetics was considering whether it could play some role in this project. At the same time, he informed Allen of Lederberg's support for citation indexes[120]. Allen immediately informed Lederberg that he had been talking about citation indexes to human geneticists "for several years without striking a spark of interest except in Everett Dempster." Maybe microbial genetics, Lederberg's specialty, was better suited?[121] Allen also invited Garfield to come over and discuss the problems of a citation index in more detail[122]. Garfield meanwhile tried to interest the Air Force in a citation index[123]. Lederberg wrote Allen that he thought that it might be better to start with a block of journals, instead of a defined area of research[124]. Allen agreed with this, as did Garfield[125]. Garfield also got the idea of enrolling scientists themselves in compiling the index:

---

[115] "I would make a distinction between the "dolts" you deal with in the Patent Office and in NSF, AEC etc. I don't know the Information specialists, but have the highest regard for the research-grant people in all these agencies" (Lederberg to Garfield, June 26, 1959).

[116] Garfield to Lederberg, July 6, 1959.

[117] "I can well see the shortcomings of the proposal, particularly in retrospect." (Garfield to Lederberg, July 6, 1959)

[118] Garfield refers to the Biology Division Library of the Food & Agriculture Organization, which maintained a citation index in card form (Garfield to Allen, April 11, 1959).

[119] Garfield to Gray, July 6, 1959.

[120] The two geneticists had known each other for a long time but had been unaware of their mutual interest in citation indexes. Garfield informed Lederberg of Allen's involvement in his first reply to Lederberg's initial letter. Lederberg wrote him that he had known Allen "for a long time" but had never discussed citation indexes with him. July 6, 1959, Garfield wrote to Allen and informed him of Lederberg's interest. By that time, the two scientists had not yet discussed the matter with each other. (Garfield to Allen, July 6, 1959; Garfield to Lederberg, May 21, 1959; Lederberg to Garfield, June 6, 1959).

[121] Allen to Lederberg, July 8, 1959; Allen to Garfield, July 8, 1959.

[122] Allen to Garfield, July 13, 1959.

[123] Garfield to Allen, July 15, 1959.

[124] Allen to Garfield, July 17, 1959.

[125] Garfield to Allen, July 20, 1959.

> I intend to write a series of letters to librarians, literature scientists and laborary scientists, suggesting that they volunteer their time for Citation Index work. He would be assigned a particular journal to work on. After all of the file cards had been sorted by journal, we could then turn over to the editor of each journal an individual Citation Index for his journal.[126]

NSF evaded committing itself and in a preliminary response[127] proposed that the geneticists themselves should take the initiative to form an advisory committee[128]. Therefore they approached NIH as well. Lederberg talked to Katherine Wilson, executive secretary of the Genetics Study Section at NIH[129], and so later on did Allen[130]. Lederberg moreover exerted pressure on NSF to act[131]. By now they were both lobbying as well as developing practical solutions to the many problems of compiling a working citation index[132]. An important issue was where to start. Lederberg had his doubts about the choice of genetics:

> Frankly I don't think that "genetics" is the first best circumscribed field for a tryout in C.I. I think "physiology" would be much better. But we need it in all of science and any move to get started is a good one. If you C.I. just a group of genetics journals as a closed group I'm afraid you wouldn't generate much improvement over what most people in the field remember from their own experience.[133]

Garfield felt the same way. He preferred to start with general science journals like *Nature*. This would, however, exacerbate the problem of the size of the index. Having each scientific journal publish a citation index as a yearly supplement might be a solution[134].

---

[126] Garfield to Allen, July 20, 1959.

[127] Program director Gray was in Europe when Garfield's letter of July 6 arrived, so the matter was handled by Assistant Program Director Tolkan.

[128] "I am not sure that it would be appropriate for the Foundation to organize a committee of geneticists to advise on a citation index project in the field of genetics. Rather, if geneticists are interested, I should think they would more properly and more effectively take the initiative in setting up such a committee, which might function through the American Society of Human Genetics or the Genetics Society of America, or at least have the endorsement of one of the societies." (Tolkan to Garfield, July 16, 1959)

[129] Lederberg had just resigned from this panel — which had triggered the idea of contacting Garfield in the first place — so he could not "nurse the idea along" as well as he could have, he wrote Garfield (Lederberg to Garfield, July 9, 1959).

[130] Allen to Garfield, July 31, 1959.

[131] All three felt that they had to overcome stubborn resistance, overcoming of which required careful tactics. "NSF may have deflected our ploy on CI but I am still thinking what might be done. Let me know if there are any further developments—there should be!" (Lederberg to Garfield, undated, between July 14 and August 3, 1959).

[132] For example, Lederberg sent Garfield some of his published work, as samples to work from, and got reprimanded by Garfield about his imprecise way of referring to other researchers: "citations ought to be a little more specific. I don't understand why citations frequently aren't more specific as to page number." (Garfield to Lederberg, July 14, 1959).

[133] Lederberg to Garfield, undated, between July 14 and August 3, 1959.

[134] "I got a good idea, I think, for overcoming the problem of size of a CI. Why can't each journal issue a yearly supplement, just as they do their index, which would be a citation index for that year?" (Garfield to Lederberg, August 3, 1959)

In August, Garfield discussed the citation index at NSF and NIH. He also had a thorough discussion with Allen about how to show the power of citation indexing: "To really demonstrate the value of a citation index we should, somehow, come up with as complete a citation index as possible to a selected list of journals and/or articles"[135]. Garfield and Allen concluded that "some mechanical method" should be developed for copying the citations[136]. The main area for intellectual problems would be the specification of the "kind of citation". Garfield hoped that the citation index would help standardize scientists' referencing behaviour:

> I believe that citation index research will pay off handsomely in the future in that this research will characterize all the different ways in which people "cite" the earlier literature. We will then be able to provide editors with a guide to standardized citation practices. Further, they might be influenced to adopt a notation or terminology that would indicate to the reader and the bibliographer the "nature" of the citation. In this project we would try to characterize, for a selected list of articles, each citation as to whether it was: 1. Review article (Rev.); 2. Communication (Com.); 3. Editorial (Edit.); 4. Errata (Err.); 5. Translation (Tr.); 6. Abstract (Ab.); 7. Book (Bk.); 8. Discussion (Disc.); 9. Summary (Summ.); 10. Bibliography (Bibl.); Book Review. I have purposely left out: refutation, confirmation, etc. I have also left out any mention as to whether pertinent portion of citing paper is experimental, theoretical, introductory or whether it is a use of a method cited or use of "material" cited. These are points to be investigated later.[137]

He also preferred to indicate page numbers to "speed up locating the pertinent statements". After his discussions with Allen, Garfield personally ran some tests on samples to discover the characteristics of the references, the average number of references in an article, and the speed with which a citation index could be compiled:

> Scanning 1500 articles took about 15 hours in five sessions of three hours duration. I could sometimes scan as much as 200 articles per hour. It was never lower than 100 per hour. Depending upon motivation skilled clerks could scan at an average rate of about 100 articles per hour.[138]

His conclusions were that the citation index was indeed feasible[139]. From the tests, Garfield concluded that the project should be a three part program:

[135]"Compiling a citation index to a selected list of articles would increase problem of scanning the bibliographies and references in articles from which citations would be taken. For example, if a paper in "Nature" is included in our sample, we would have to carefully examine citations to Nature. However, this would offset the cost of handling a larger quantity of citations in non-genetics articles." (Garfield to Lederberg, September 9, 1959, originally prepared August 15, 1959).

[136]Two years before, Garfield had discussed these matters with the National Library of Medicine, which could build a special-purpose camera for about $ 1,000.

[137]Garfield to Lederberg, September 9, 1959 (originally prepared August 15, 1959).

[138]Garfield to Lederberg, September 3, 1959.

[139]"The conclusion to be drawn from test#2 is that we can compile a complete citation index to all the general science journals, making the sample not only interdisciplinary but permanently useful when it is finished. The work will not be wasted. This ties in beautifully with another idea I had

> In conclusion, pending comments from you, G. Allen and others[140], I feel that a revised proposal to NSF should be based on the following three part program of Citation Index research.
>
> 1. Mechanically (photographically) pick up <u>all</u> references found in a specified list of genetics journals <u>and</u> articles, the latter based on some well known genetics bibliography. From these eliminate undesired references.
> 2. Scan all Current Content journals for references <u>to</u> all articles appearing in a <u>specified</u> list of genetics journals <u>and</u> a specified list of articles or authors.
> 3. Scan a <u>large</u> list of journals from <u>all</u> representative scientific disciplines for references <u>to</u> general science journals including Nature, Science, Proc. Natl. Acad., etc.[141]

This would result, according to Garfield, in three different indexes: a complete citation index of a list of genetics journals, a complete index of a list of articles in non-genetics journals and an index of all articles in general science journals. Garfield wished to scan "at least the last five years of the literature, preferably more". They could be published as "individual journal articles or supplements", or as a single combined "Citation Index to Genetics". Testing should be done by handing out copies of the index to "geneticists and various libraries" and awaiting their comments. Lederberg "strongly disagreed" with "the addition of interpretative material" such as classifying the type of references or subjects:

> I also disagree that you should attempt any subject classification, e.g., genetics, on the grounds that this defeats the main advantageous purposes of CI, namely a *mechanical* system of classification. If you start analysing the references, you might as well start trying to analyse the content of the paper, and you are back to abstracts.[142]

He also emphasized the necessity of differing from the traditional, disciplinary approaches: "a good general CI will be of greater value to Genetics than a too specialized run that sticks too closely to the discipline". He therefore preferred the citation index to the general science journals, Garfield's research program's third part. Contrary to Garfield, Lederberg had the strong feeling that there was actually no further research needed:

> If you stick to your guns on the original principles of CI, I am sure you will find it widely used as a research tool, and that further perfections will evolve.

---

and discussed with Gordon Allen in which we would abandon the concept of a unified citation index to all science journals and prepare, instead, <u>individual</u> Citation Index for <u>each</u> journal. At the end of each year we could send to each journal editor a citation index for his own journal. Periodically the individual Citation Indexes could be accumulated. This would be similar to the practices followed for legal citation indexes. (Garfield to Lederberg, September 3, 1959)

[140] Garfield sent a copy of his letter to Gordon Allen, Katherine Wilson (NIH), and Connie Tolkan and George LeFevre (NSF).

[141] Garfield to Lederberg, September 3, 1959.

[142] Lederberg to Garfield, October 6, 1959.

> What we need more than anything else is to get it going. (...) I know that you yourself will be keener to do the kind of analysis you're discussing than just to go ahead with CI as is. But I think there will be much more support for you if you can demonstrate what CI can do. (...) My main aim, as you know, is to encourage you to get on with the work as simply and straightforwardly as possible. It it works out as well as it must, you should have little concern for enthusiastic support for your own research using CI.[143]

2.3.7 Growing support

The idea of citation indexing scientific literature started to catch on slowly. The board of the American Society for Human Genetics reacted sympathetically[144]. NIH's Katherine Wilson urged Gordon Allen not to wait for action from her group but proceed with their plans[145]. At NSF, the mood was changing. "Things have changed, in that they are now listening very carefully because you and the genetics people have shown some interest", reported Garfield[146]. He had even been told that he did not have to stick to $ 29,000 per year, the amount of money did not matter provided the idea was accepted[147]. The foundation insisted on a committee of scientists, however, and was not willing to take up Lederberg's idea to form it itself[148]. The Human Genetics Society was also reluctant, so Garfield started phoning researchers himself[149].

Lederberg paid him his first visit in Philadelphia the same month[150]. They discussed the committee in more detail. Lederberg had by now developed firm opinions on the way the project should be organized. As Garfield reported a few days later to Gordon Allen[151], Lederberg stressed that the committee should not meet too often[152], that the project should run for at least three years and that it

[143]Lederberg to Garfield, October 6, 1959.

[144]The board had met August 30, 1959 at State College, Pennsylvania. Gordon Allen explained the principles of citation indexing as a scheduled item on the agenda. None of the board members present expressed doubt of the potential value of a citation index, Allen reported to Garfield and NSF director Dwight Gray. The society did not, however, see it fit to take action as society. Allen got permission to report "their favorable response as individuals". (Allen to Garfield, October 26, 1959)

[145]Allen to Garfield, October 6, 1959.

[146]Garfield to Lederberg, October 9, 1959.

[147]Garfield to Lederberg, October 9, 1959.

[148]Garfield to Lederberg, October 12, 1959.

[149]By that time, Garfield only had the support of Allen and Lederberg, with the added practical problem that Lederberg worked at Stanford and Allen in Washington. Therefore, he wished to concentrate the committee either in the West or in the East of the States. He asked Lederberg to get "two or three nearby in or out of genetics". He clearly counted on Lederberg: "Finally, if I am being presumptuous in assuming you can work on this committee at all then let me know that too, which admittedly would be a disappointment, but I've gotten used to that. In that case we would again select all Easterners" (Garfield to Lederberg, October 12, 1959).

[150]This meeting took place on the Friday before October 29, 1959, and was the first time the two men met (Joshua Lederberg, Interview, February 3, 1992, New York; Garfield to Allen, October 29, 1959).

[151]Garfield to Allen, October 29, 1959.

[152]He was willing to attend one meeting per year. An optional "local committee", comprised by Garfield, Allen and people from the East coast, could meet more often.

should process literature produced during the previous five years. He further urged Garfield to apply both to NSF and NIH. This would enable coverage of all types of journals. Garfield wholeheartedly agreed to this: "Since the file that would result from this would enable us in the future to do anything but attack the literature comprehensively"[153]. This was underlined by their conclusion that "it costs only about one cent per citation to prepare a punched card for subsequent use in a computer". Hence, it would be no more expensive than selectively entering citations to a particular set of journals. Their discussion did not turn up anything really new, but it did lay the ground plan for the citation index project as it would emerge later. Stimulated by NSF's remark that money was not the main problem, they increased their budget to around $ 200,000 for an eight-year period. This meeting was also the first time they discussed the idea of a national newspaper for science[154]. Later, Garfield attended a meeting at NIH where he presented a dummy of a tabloid size newspaper. This newspaper would have listings of science communications as well as citation indexes to them.

2.3.8 Delay

Although in October 1959, Garfield intended to submit the proposal "in the very near future"[155], it actually took a little over five months. One reason for this was that another proposal Garfield had written at the request of NIH was subsequently turned down. Once again, the profit-making status of his company proved an important obstacle[156]. Not only did this make Garfield wonder whether his proposals had any chance at all[157], it also stimulated him in changing the organization of his work. He contemplated the formation of a non-profit institute but was advised not to do so[158], by amongst others Chauncy Leake[159]. He did, however, settle on a new name to bestow a more "acceptable" status — the Institute for Scientific Information[160], although this took time[161]. Garfield was moreover involved in "a lot of work" because of the production and increased acceptance of *Current Contents*[162]: "This has slowed down somewhat my prepa-

[153] Garfield to Allen, October 29, 1959.

[154] Garfield to Lederberg, November 17, 1959.

[155] Garfield to Allen, October 29, 1959.

[156] Garfield to Lederberg, October 12, 1959.

[157] "Since it takes a lot of time, money, and energy to prepare these proposals I am not anxious to work them up if we are beat before we start."

[158] The main reason was a possible conflict of interest between Garfield Associates and its affiliated non-profit institute.

[159] Garfield to Lederberg, October 12, 1959.

[160] "In this way we would have the benefit of an "acceptable" name and none of the legalistic difficulties of forming a non-profit institute." (Garfield to Lederberg, October 12, 1959)

[161] In March 1960, Garfield explained the delay in a letter to the members of the prospective advisory committee Lederberg, Allen, LeFevre, Melnick, and Spiegelman: "Perhaps you have wondered why I have not written sooner concerning our plans on Citation Index research. This letter is meant to advise you that I have not given up the idea of pursuing this project. However, recent developments have necessitated a brief delay in forwarding our revised proposal to the National Science Foundation." (Garfield to Lederberg, Allen, LeFevre, Melnick and Spiegelman, March 14, 1960).

[162] Garfield to Lederberg, March 14, 1960.

ration of a proposal for citation index research."[163] In December 1959, he was told by NSF that it was highly improbable that funds would be available from the year's fiscal budget "as a result of heavy grants to Chemical Abstracts and Western Reserve Univ."[164]. Garfield decided to defer submission of the citation index proposal to "on or before March 1st". Moreover, Garfield had written several other proposals to NSF which also needed funding. He had submitted a grant proposal for a *Current Contents* edition for space and physical sciences, as well as a proposal for another major project, the *Index Chemicus*. The latter project was one of Garfield's top priorities[165]. Garfield was also working on a "Copywriter", a device which would, Garfield hoped, also help to compile a citation index[166].

The citation project was delayed but had not completely ground to a halt. Garfield continued computing various ways of compiling the citation index, focusing on the comparison of comprehensively processing all citations with the selection of citations to a journal like *Science*[167]. He also paid renewed attention to the problem of publishing a future citation index. This did not prevent Lederberg from becoming a little worried. He inquired whether Garfield had considered creating the possibility for private investors to buy stock in his new Institute for Scientific Information[168]. This question was motivated by "some anxiety about getting Citation Indexing under way"[169], Lederberg explained. He understood Garfield's "being fed up with the federal grants situation" and proposed two possible other avenues: getting help from private foundations (notably the Rockefeller Foundation) and "raising capital by public subscription",

> which I bet you could do on SCI* alone, apart from your other important contributions. It is refreshing to see the kind of action you have generated while everyone else is talking, and while we have disagreed on some minor details, you certainly do have my confidence in the way indicated by my inquiry.
> ∗ SCI = Science Citation Index (free gift to you).[170]

Note that Joshua Lederberg coined the now famous name *Science Citation Index*, which he especially liked because the abbreviation *SCI* stresses the link with science[171]. He stressed the importance of the citation index, which in his opinion

163 Garfield to Lederberg, November 17, 1959.

164 Garfield to Lederberg, December 14, 1959.

165 Garfield to Lederberg, March 14, 1960.

166 Lederberg suggested two private firms to Garfield that could help develop his Copywriter (Lederberg to Garfield, received April 4, 1960).

167 Garfield to Lederberg, November 17, 1959.

168 "As a query in passing, is there an opportunity for private investment in your new Institute for Scientific Information? I might be quite interested myself; if you were interested further I suspect some of my colleagues would be too, though I am sure you do not lack very much in this direction." (Lederberg to Garfield, April 8, 1959).

169 Lederberg to Garfield, April 23, 1960.

170 Lederberg had some recent experience when he wrote this. The Rockefeller Foundation had supported his planetary microscope evaluation project "knowing that NASA was embroiled in the usual government red tape". (Lederberg to Garfield, April 23, 1960) Garfield accepted Lederberg's idea to call the index the *SCI* rather off-handedly: "Incidentally, thanks for the gift". (Garfield to Lederberg, April 26, 1960)

171 Joshua Lederberg, Personal Interview, February 3, 1992, New York.

"beats the other projects hands down in importance". Garfield tried to reassure Lederberg:

> As far as the relative importance of the various projects is concerned, I agree with you wholeheartedly that SCI is at the top of the list. That is why I want to put my best foot forward on it. I don't want to mess it up by careless planning and thinking and, as you know, there are definitely some problems in carrying it out.[172]

He wished moreover to stick to trying to get a grant from NSF, not believing in the stock selling potential of the *SCI*, nor having had encouraging experiences with the Rockefeller Foundation[173].

2.3.9 Submission

One month later, Garfield sent in his proposal to NSF and asked Lederberg to support it[174]. Lederberg promised to do so[175], and wrote the same day to NSF in no uncertain terms:

> I am sure there would be little point in adding to the testimony of scientific exasperation at the tremendous problem of coping with the existing scientific literature. There is no one solution to this problem; albeit the topical abstracting services do perform a useful function. To my own mind, and this is a considered conclusion, the Citation Index would be of inestimable value in improving the efficiency of scientific research insofar as this depends on useful access to the literature. (...) I am deeply and enthusiastically interested in the success of Mr Garfield's proposal and am happy to have the opportunity to support this endeavor by serving on his advisory committee.[176]

Lederberg stressed *SCI*'s "special value in interdisciplinary work", especially if it would be compiled "on a global scale so as to encompass the entire literature":

> If this problem were entirely in my own hands, and this is the main point on which I differ with Mr. Garfield, I would not confine the initial efforts to a single field but would waste as little time as possible in securing comprehensive coverage.[177].

---

[172]Garfield to Lederberg, April 26, 1960.

[173]"I have tried to work with the Rockefeller Foundation on other matters, such as Current Contents, and drew a complete blank. Generally they are not terribly interested in problems related to scientific documentation. (Garfield to Lederberg, April 26, 1960).

[174]"A call to either of these fellows would not hurt. The last time you said something to somebody about citation indexes things started to happen." (Garfield to Lederberg, May 23, 1960).

[175]"As you suggested I will write to NSF myself to verify my own specific interest in the program. I have in mind also communicating with George Kistiakowsky, Scientific Adviser to the President, but it may be better to wait to see what happens at NSF as this would amount to going over their heads" (Lederberg to Garfield, May 27, 1960).

[176]Lederberg to Adkinson, May 27, 1960.

[177]Lederberg also expressed this preference to Garfield: "I think the proposal for a separate Citation Index to each journal is an ingenious one but really a stopgap. The full program, in my opinion, would call for an independent publication of the unified SCI" (Lederberg to Garfield, May 27, 1960).

Lederberg found Garfield's proposal "quite impressive"[178]. Allen nevertheless proposed "a radical change in the design". In the draft proposal, Garfield opted for a selective scanning of citations to genetics journals. Allen agreed with Lederberg that they should "not fool with the selective scanning, but go immediately to full coverage for a trial period": "I still feel that, if the necessary budget and facilities could be obtained, complete coverage would be a better approach"[179]. He also provided Garfield with a set of biochemistry articles that proved the usefulness of the *SCI*[180] and gave him advice on the precise formulation of several crucial passages in the proposal[181]. Garfield changed the proposal in accordance with Allen's comments and proposed gently steering towards a comprehensive compilation:

> I am glad that you agree that the comprehensive approach would be more satisfactory. My plan has been this—to wait for NSF's reaction, but no later than two weeks from now—and then send in a proposal to NIH and in this proposal we will go for the comprehensive approach and sacrifice the number of journals covered—particularly in the physical scienes—so that we can have a low enough budget to do what we do completely. (...) I don't think there is any opposition from our committee to the comprehensive approach. It is a question of how to convince the NSF people.[182]

[178]Lederberg to Garfield, May 27, 1960. The other members of the committee reacted favourably as well. LeFevre thought it was okay (Garfield to Lederberg, June 1, 1960), while Allen wrote that the application looked "very good" (Allen to Garfield, May 27, 1960).

[179]Allen emphasized this in a renewed statement of his position: "If I am in any way responsible for the selective scanning idea and concentration on genetics, I now wish to reverse myself. If others on the committee express this same view, perhaps you should consider reducing the number of journals scanned, and reading out all the citations in those journals. This would be more accurate, would really cut down the labor of compiling a more extensive SCI in the future, and would still permit you to sort out the journals cited to permit compilation of citations in genetics and general science". (Allen to Garfield, May 27, 1960).

[180]"each one appeared, when published, to be correct and final, but each has been subsequently disproved or significantly extended. A person who read one of the earlier articles would be unlikely to suspect its inadequacies, and without a citation index at his elbow would probably proceed in ignorance of the current state of knowledge." (Allen to Garfield, May 27, 1960)

[181]He advised Garfield to revise his statement about the choice of genetics as the field for the citation index: "you singled out genetics as a discipline differing from others in its indexing requirements. Perhaps I did once say that genetics was particularly in need of a citation index, but what I think I meant, and certainly what I would say now, is that many illustrations of the need can be found in genetics. I believe that every area of active scientific advance would benefit about as much as genetics." He moreover showed Garfield that the mathematical relationship between the number of references in an article and the number of citations was far from straightforward, mainly because science is not in a "perfect steady state but it is growing and expanding rapidly" (Allen to Garfield, May 27, 1960). This was not only a theoretical but a practical question as well, because the number of citations to be processed would determine the compilation costs. Garfield thought about this himself a lot since he had found that the average number of references per paper in the biological literature was 15. He concluded that he did not have to be very worried about the average length of citation index entry. Garfield double-checked the labour involved by having the bookkeeper type in citations for two hours. (Garfield to Allen, June 1, 1960, and June 3, 1960).

[182]Garfield to Allen, June 1, 1960.

One day later, Allen sent his set of biochemistry articles together with a drawing of the citation relationships between them. Gordon Allen had, in other words, drawn the first citation network:

> The arrows indicate the direction in which one would be led in a conventional literature search, starting at any point on the network. A citation index would permit one to trace the arrows in the opposite direction, and hence to find all the articles no matter where on the network he started.[183]

He emphasized that this small network was an extract from "a considerably more voluminous literature on the same topic, all tied together with citations". Garfield reacted strongly:

> The material you sent me is magnificent! This must have been a great deal of work. It is fabulous. Why didn't we think to do this before. I didn't have this in mind when I said I had some examples of the power of the Citation Index. I merely meant specific articles which could be traced through a CI. (...) I once had the idea that some type of network theory could be used with Citation Indexes. I am now convinced more than ever, from your example, that this will be true.

NSF's first reaction to Garfield's submission[184] —in a letter to Lederberg— clearly demonstrated its ambiguous attitude towards citation indexing:

> This technique is one of a number about which people long have had a variety of subjective opinions — some pro and some con — without there having been any really sound experimental investigation of the idea. The citation index approach would seem to have considerable promise for leading the researcher rapidly and efficiently to the significant literature in a given subject area with, perhaps, the major element of possible weakness being the fact that its effectiveness necessarily depends on the care and seriousness with which authors of papers select the references they cite.[185]

This attitude was "not altogether negative"[186], though the reaction annoyed Lederberg who began to understand Garfield's impatience:

> In fact, I am beginning to get a glimmer of understanding as to the basis of and the magnitude of the hostility or misappreciation for SCI. Just because it is a tool which will be handled automatically by scientists themselves it will tend to give less scope and importance to professional information-handling bureaucracy.[187]

[183] Allen to Garfield, June 2, 1960.

[184] NSF had also been approached by Allen who, like Lederberg, indicated his preference for a comprehensive citation index (Allen to Adkinson, June 6, 1960).

[185] Fry to Lederberg, June 3, 1960.

[186] Lederberg to Garfield, June 21, 1960. Garfield did not fully trust this: "sounds promising but so did it the last time I sent in a proposal. They never discourage you before you do all the work!", he wrote to Allen (Garfield to Allen, June 24, 1960).

[187] Lederberg to Garfield, June 21, 1960.

Garfield could not agree more:

> I am glad that you are independently getting a glimmer of the kind of hostility I have encountered—not only with regard to the citation index, but with regard to any ideas which tend to free the scientist from dependence on an intermediary—whether it be the librarian, the government bureaucrat or what have you.[188]

On June 9, Garfield got a telephone call from NSF, telling him that it would take at least four if not five, months before NSF could make a decision[189]. Two weeks later, Garfield sent his proposal to NIH (Garfield 1960*b*), which was "almost the same as to NSF except I added a modification suggesting the comprehensive approach"[190]. He also included Allen's citation network sample as an appendix[191]. By now, Garfield had actually decided to go on with the citation index regardless of the reactions of NSF and NIH:

> I will let nothing stand in the way of getting a citation index going. (...) The more I think about C.I. the closer I get to agreeing with you that it might be turned into a practical ISI project regardless what NSF and NIH decide.[192]

Moreover, he moved to form a non-profit organization[193] "to take over the work of the SCI project"[194]. On July 7, NSF phoned Garfield to suggest some minor changes. Garfield included most of them[195] and resubmitted the proposal on July 15 (Garfield 1960*a*)[196].

[188] Garfield to Lederberg, June 24, 1960.

[189] Reasons given were the summer vacation and the end of the fiscal year (Garfield to Lederberg, June 9, 1960).

[190] Garfield to Lederberg, Jun 24, 1960. Garfield included "almost all" changes Allen had recommended (Garfield to Allen, June 24, 1960).

[191] He also showed them to Lederberg: "I thought you would appreciate seeing this rather interesting "network" diagram G. Allen worked out (Garfield to Lederberg, June 24, 1960).

[192] The latter part of the quote is pencilled in ink, after Garfield's secretary had already typed the letter (Garfield to Lederberg, June 24, 1960).

[193] This was partly motivated to prevent conflict of interest situations. An example of this surfaced in a discussion between Allen and Garfield about Allen's position in NIH. Allen had still not heard from NSF when he informed Garfield that his name on the application to NIH might "carry bad luck". He had "a strong impression" that Walter Burdette, chairman of the NIH Genetics Study Section, mistrusted or disliked him and that he might be accused of a conflict of interest (Allen to Garfield, June 28, 1960 and July 11, 1960). Garfield did not like the idea of deleting Allen's name at all: "Unless you give me some strong reasons for eliminating your name I see no reason to do so. Naturally I don't want to jeopordize (sic) the project but I also don't intend to desert my friends for a few sheckels. Let the truth be known. You were one of the first to take up my proposal and I am grateful to you for it and for the great deal of time you have put into it (Garfield to Allen, July 7, 1960). Garfield later reassured Allen, informing him of the formation of a non-profit organization: "I suppose one could still have a conflict of interest regarding a non-profit organization but I suppose that it would be harder for anyone to say it was for "profit" reasons." (Garfield to Allen, July 26, 1960) Allen complied: "Removing my name from the application would not hurt me in any way, and might help you, but perhaps it is best to reveal my association with the idea, for better or for worse. As for myself, I am proud of the association. (Allen to Garfield, July 11, 1960)

[194] Garfield to Allen, July 26, 1960.

[195] Garfield to Allen, July 7, 1960.

[196] NSF acknowledged its reception July 21, 1960 (Gray to Garfield, July 21, 1960).

2.3.10 The genetics proposals

The proposal for a genetics citation index to NSF was thoroughly rewritten and no longer comparable with the feasibility study applications of 1958. Bringing the scientist up-to-date on a particular paper was the first application mentioned. The overcoming of "artificial dividing lines" between existing abstracting services was mentioned as "a special advantage" of the *SCI*. The proposal moreover stressed that it was not meant to replace existing subject indexes and that the *SCI* would be especially useful for writers of review articles. These applications were supported by Gordon Allen's citation network (figure 2.2 on page 54) which was added in the appendix[197]. Garfield made clear that "one of the most attractive features" was the susceptibility of the index to "complete mechanization":

> Compilation by a staff of trained scientists is not necessary in order to index papers as the "indexing" has already been done by authors in providing citations to earlier papers. Compiling the *SCI* is almost completely a routine task of copying citations in new papers, sorting them in order by journal, year and page (so that all references to the same paper will be brought together), and then distributing the information either as a printed bibliography or in card form. (Garfield 1960*a*)

This argument was reinforced with the tabulations and testing Garfield had undertaken during the previous months:

> The primary factor determing (sic) the feasibility of compiling a *SCI* is a quantitative one. The first general impression is that there are so many references in the literature as to make a *SCI* huge and unwieldy. Fortunately this is not true, as extensive preliminary studies have shown.

These computations relied heavily on Garfield's experiences with *Current Contents* which by this time processed around 600 journals[198]. Sixty per cent of the journals needed for the genetics citation index were already in house thanks to *Current Contents*. Garfield based his reasoning on his finding that the average number of references per article in the biological literature was fifteen. The CC-journals published 125,000 articles per year. Extending coverage to 1000 journals would lead to three million citations per year. Based on the test runs he expected production costs woule be two to three cents per citation, i.e. between sixty and ninety thousand dollars per year. Garfield found it a surprisingly low figure when compared with the several million dollars needed to abstract and index the same amount of articles in the conventional way. He subsequently showed

[197] The appendices of the proposal to NSF (Garfield 1960*a*) consisted of: Gordon Allen's sample of biochemistry references; the diagram of its citation network; the actual appearance of a printed citation index of Allen's sample; the physical science journals covered by *Current Contents*; the life science journals covered by *Current Contents*; Garfield (1955); Adair (1955); Garfield (1957); Seidel (1949); Hart (1949); Garfield (1956*a*). The NSF proposal and its first three appendices were added as appendices to the NIH proposal (Garfield 1960*b*), while 40 copies of the articles on citation indexing were separately attached for the reviewers.

[198] An alphabetical list of the life science journals covered (slightly more than 500) was added to the proposal as an appendix (Garfield 1960*a*).

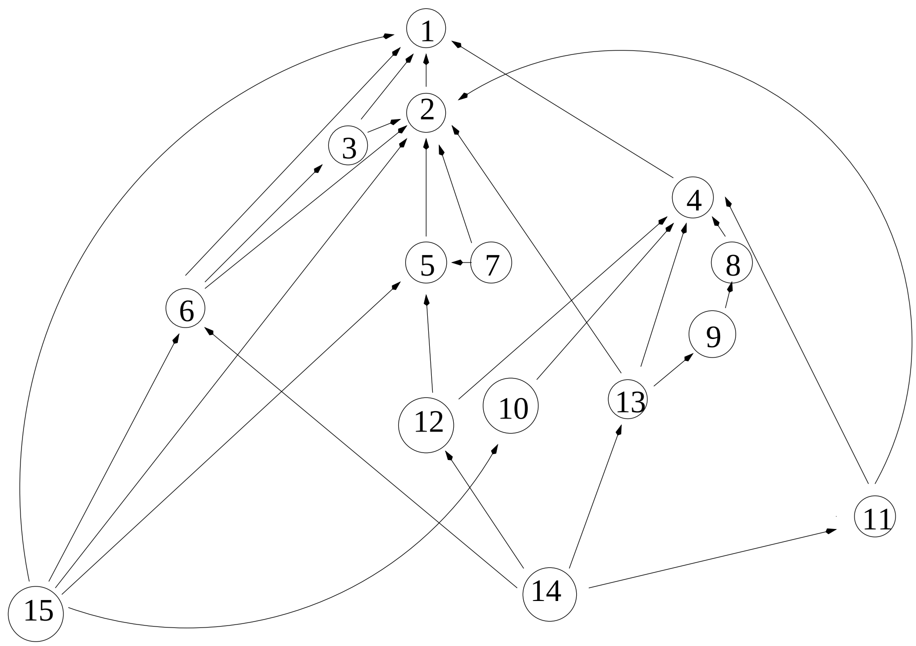

Figure 2.2: Gordon Allen's citation network as depicted in Garfield (1960*a*). The circled numbers represent published articles. The arrows indicate citing relations, pointing from the citing to the cited document.

that selective compilation could actually lower the number of processed citations while at the same time its production per citation would be four times more expensive. Therefore:

> As the number of selection criteria increases and the number of pertinent references found increases one rapidly reaches the point where it is cheaper and more efficient to process every citation. (Garfield 1960*a*, 4)

In the proposal to NIH, Garfield was even more explicit:

> Allen and Lederberg (and myself) are of the opinion that the comprehensive approach is more sensible, economical and productive. If I were in a position to ask for the ideal support I would ask for sufficient funds to cover the compilation of a complete Science Citation Index (SCI), covering both current and old literature, utilizing evaluative and editorial techniques for building the Citation Index into a true encyclopedia of unified science. (Garfield 1960*b*, 5)

In the face of "the immediate practical difficulties of obtaining such unlimited support", he sketched the compilation of "a body of information which could be added to without difficulty if further additional support were found later" (Garfield 1960*b*, 5). In his proposal to NSF, Garfield proposed to follow the selective approach "in order to keep the budget of this research project as low as possible". The money thereby saved could be used to process a five year backlog which was necessary to demonstrate the value of the citation index. This resulted in the proposal to NSF (Garfield 1960*a*) to construct a citation index by scanning a

J. AMER. CHEM. SOC. 63, 69 (1941) #1
Michaelis,L. Cold. Sp. Harb.Symp.XII,131(1947) (#2)
Michaelis,L. J. Phys. Coll. Chem. 54,1(1950) (#3)
Zanker, V. Z. Physik. Chem. 199,225(1952) (#4)
Lawley, P.D. Biochim. Biophys.Acta 19,328(1956)(#6)
Schoenberg,M.D. in press (#15)

COLD SPRING HARBOR. SYMP. XII,131(1947) #2
Michaelis,L. J. Phys. Coll. Chem. 54,1(1950) (#3)

Figure 2.3: First lines of the example of the actual appearance of a printed *SCI* as included in the proposals to NIH and NSF.

list of 1000 journals and processing all references to 43 specified genetics journals as well as all references to 22 specified general science journals. The proposal to NIH, on the contrary, entailed the processing of *all* references from the specified journals, punching them into IBM cards which were subsequently sorted mechanically. Only then would the citations to genetics journals be selected and printed (Garfield 1960*b*)[199]. If both NSF and NIH were to decide to support the project, the NSF funds would be used to "cover backlog" (i.e. process older literature) and the NIH money could be directed to increase the coverage of current literature[200].

The index would be published by providing the editors of the journals with an individual journal citation index which could then be published as a yearly supplement. This was only an intermeditate mechanism, though. Garfield wished to keep the option of a separate publication open and wrote that he was "in correspondence with editors on the publication problem". As figure 2.3 on page 55 shows, Garfield was by now thinking of printing the *SCI* in bibliography style. The total budget was estimated at $ 156,000 for three years[201].

2.3.11 Convincing NSF and NIH

In order to win over NSF, Garfield prepared the formation of a non-profit organization, "The National Documentation Center". He first sounded out Lederberg about this, inquiring whether this would lessen his interest in future ISI stock[202].

[199] Garfield did not conceal this difference: "This method stands in sharp contrast to an alternative method proposed in a co-pending application to the National Science Foundation." He explained this by referring to the existing "dichotomy of opinion on how to approach the problem of studying citation indexes" (Garfield 1960*b*).

[200] "Essentially the same proposal has been submitted to the National Science Foundation. If support were received from both NSF and NIH then each grant would be used to support additional research. The NSF funds would be used to cover backlog, NIH to cover current literature." (Garfield 1960*b*, 3)

[201] Eight scanner-key punchers, "if feasible part-time graduate library students", would produce the punched cards. Part-time project supervisor would be Gwen Bedford, University of Pennsylvania, who had also written the 1958 feasibility study proposals.

[202] Garfield to Lederberg, July 22, 1960.

The geneticist agreed that "if this will help to secure and maintain federal backing it would be desirable to entrust the development of SCI to a non-profit organization"[203]. He was less certain about the name: "it has some connotation of a governmental activity and if this is felt and thought to be presumptuous, it might do more harm than good"[204]. When Garfield informed NIH of the NDC plan, secretary Wilson reacted immediately, telling Garfield that this would have "an important impact at NIH"[205]. By this time, Garfield had experienced the effects of the new name of his institute:

> You would be amazed (...) how many doors the new name has opened for us. There certainly is something in a name!! I have never seen anything like it.[206]

Nevertheless, Garfield did not really like the non-profit angle and meant to pursue this route only if it was necessary to convince NSF and NIH.

Lederberg, thinking about how to get NIH support, suggested that Garfield would bring up the advantage of the *SCI* in evaluating the impact of funding with Katherine Wilson:

> Quite seriously with so many agencies anxious to know just what their real effect is, a quantitative measure such as SCI would very readily furnish would be a very valuable tool for them.[207]

Garfield agreed, telling Lederberg that in the past he had tried to sell *SCI* to the Air Force "for exactly the reason given in your letter"[208].

At NIH, discussion moved to the issue of publishing the *SCI*[209]. November 7, NSF resumed contact with Garfield[210], explaining that the delay had been partly caused by the similarity of both of his proposals to NSF and NIH[211]. The agency

[203] Lederberg to Garfield, August 1, 1960.

[204] Lederberg therefore prosed to find another adjective than "national".

[205] Garfield to Lederberg, August 13, 1960.

[206] Garfield to Lederberg, August 13, 1960.

[207] Lederberg to Garfield, July 29, 1960.

[208] Garfield to Lederberg, August 13, 1960.

[209] Wilson put Garfield in contact with Burdette who asked him whether Garfield could get out a printed publication within the budget outline. Garfield got the impression that NIH wished to subsidize the compilation while he would carry the responsibility to publish the *SCI* "at a reasonable price". He concluded that a non-profit organization might not be needed but was not sure. Burdette also asked Garfield about the advisory committee which made him anxious that Allen might have been right after all. Garfield did not know what to make of this (Garfield to Lederberg, September 21, 1960), but neither did Lederberg: "I don't know what to make of all the cross-currents at NIH" (Lederberg to Garfield, approximately October 7, 1960).

[210] "I am sorry it has taken so long since we saw you to get in touch with you regarding the citation index proposal. You probably have thought the train disappearing permanently with us after our dash to the station with you. (...) The delay has stemmed partly from press of other work and partly from some lag in getting together with the NIH people to pool our ideas of just what each of us would want from the overall project and who would pay for what, assuming the two agencies consideration of your proposals results in a pair of green lights." (Gray to Garfield, November 7, 1960).

[211] "The fact that the two proposals as they now stand are almost identical although not covering (and not intended to cover) mutually exclusive tasks makes them a little difficult to process." (Gray to Garfield, November 7, 1960).

asked whether Garfield could not prepare "a single, composite budget which will give a total itemized breakdown that will better lend itself to joint discussion by NIH and ourselves". NSF was primarily interested in "a sound test of the value of citation indexes as a bibliographic tool", whereas NIH's primary goal was to obtain a usable genetics citation index. Garfield consequently revised the project budget and increased it to $ 100,000 per year basing it on the approach: "all journals processed will be processed comprehensively, i.e. every citation in every article will be 'carded'"[212]. Garfield reported to Lederberg: "We are definitely making significant progress on the Citation Index Project. (...) It appears that everyone now is enthusiastic about the "comprehensive approach"[213]. Two weeks later, he estimated that the project "has better than 75 per cent chance"[214]. On December 26, he could at last break the big news to Lederberg:

> Dear Josh,
> The offical note that NIH approved our grant came in the other day[215]. This was quite nice xmas present to say the least.[216]

and to Allen:

> Dear Gordon,
> Santa Claus was very good to us. We learned that NIH approved its half of the revised budget which NSF asked me to submit based on 100,000 per year for three years.[217]

Lederberg congratulated Garfield[218]. Allen, who had not heard from him since that summer, was relieved:

> This is wonderful news. You had been silent about it so long, and I thought so long after the date when a decision should have been made, that I had about given up hope.[219]

NSF needed more time than NIH. Garfield was told that he would not hear until "after New Year". "Apparently, there is some difficulty in view of the fact that NIH is making a grant to us but NSF is negotiating a contract", concluded Garfield[220]. Only two months later was this problem solved[221]. By now, Garfield

[212]By now, the number of key punchers was increased to thirteen. The budget was also signed by Marvin Schiller, who had become Associate Director of ISI. (Garfield to Gray, December 7, 1960)

[213]Garfield to Lederberg, December 2, 1960.

[214]Garfield to Lederberg, December 15, 1960.

[215]Brewer to Garfield, December 15, 1960; received December 23, 1960. The "notification and statement of grant award" granted $ 49,450 per year for three years.

[216]Garfield to Lederberg, December 26, 1960.

[217]Garfield to Allen, December 26, 1960.

[218]"Congratulations + happy new year" (Lederberg to Garfield, undated).

[219]Allen to Garfield, December 28, 1960.

[220]Garfield to Lederberg, December 26, 1960.

[221]"It seems that the National Science Foundation bureaucrats made a real big issue out of the question of negotiating a contract with a profit-making organization thus holding up their processing or our grant. However, somebody finally decided to do some thinking and came up with a rather simple solution for getting around the law which prevents NSF from making grants to profit-making organizations. (...) NSF will presumably make a grant to NIH and then NIH will make the grant to us." (Garfield to Lederberg, February 13, 1961)

"was all set", having received his first NIH check[222], and was "trying to move ahead now as fast as possible". ISI moved to a new building and prepared to float the public stock. The idea of the formation of a non-profit National Documentation Center was shelved[223]. However, in March 1961, work had not yet begun. NSF was still in the process of transferring its half of the project budget to NIH[224], which held things up. Garfield had also been busy with other projects. He had finished his doctoral dissertation, ISI had been moving to its new building, and last but not least the prospective project supervisor, Gwenn Bedford, left Pennsylvania University for Michigan. Garfield had to lead the project himself. In May, Garfield received the signed contract from the National Science Foundation: "This means that the Citation Index project is now finally started"[225].

Back to Contents Page

[222]The first payment was $ 12,364 (Brewer to Garfield, December 15, 1960).

[223]Garfield to Lederberg, February 13, 1961. Garfield inquired tentatively whether Lederberg would wish to be member of the Board of Directors of ISI (Garfield to Lederberg, February 13, 1961). The geneticist wondered if he "could not be much more useful" if he remained a proponent without a formal status (Lederberg to Garfield, March 1, 1961). Garfield "bowed to his wisdom" (Garfield to Lederberg, March 17, 1961).

[224]Garfield to Lederberg, March 17, 1961.

[225]Garfield to Lederberg, May 17, 1961.

Back to Contents Page

Chapter 3

The building of the Science Citation Index

3.1 Building the index

I think you're making history, Gene![1]

Building the *SCI* turned out to be a bigger project than even Garfield had expected. It took more time, more money and was technically more complicated than had been foreseen in the contracts. Constructing the index was not only a huge technical endeavour but also a political enterprise. Joshua Lederberg in particular perceived the *SCI* as a possible means opening up the clogged communication channels in science. Building the index required not only extensive knowledge of library science and technical expertise but also political acumen. Due to the intense co-operation between Garfield and Lederberg, this enterprise temporarily became part of the science policy debate in the United States about the now-famous Weinberg report. The history of the *SCI* can hardly be understood without an appreciation of the technical difficulties and political dimension. Nobody really knew how a science citation index would turn out. Garfield, Lederberg, Allen and their associates all had their private visions, based on their own specialty, but it was utterly unknown what kind of organization of the literature would result from citation indexing it. Even after the funding problems had been solved, the whole project still could collapse on technical grounds. The makers of the *SCI* saw this risk but were determined to take it and prevent a failure.

3.1.1 Political design

The information crisis

As we have seen in section 2.3.5 (page 39), science policy provided the context in which Lederberg was reminded of Garfield's (1955) proposal and decided to make contact. *SCI*'s political relevance was directly related to its bibliographic

[1]Lederberg to Garfield, January 24, 1962.

properties. Yet, the connection between the *SCI* and science policy was rather tenuous. From October 1961 onwards this relationship began to change. The building of the *SCI* temporarily became intimately involved in the debate on the future of scientific information in the United States. Joshua Lederberg made the connection:

> Mainly via John Tukey, I have found myself appointed to a new committee of PSAC on Scientific Information.[2]

Lederberg did not take this appointment lightly. His assignment was to rewrite the general introduction to the report and he saw this as an opportunity to push for a radical overhaul of the "anarchic" way scientific information still was organized. Derek Price's *Science since Babylon* had been one of the stimulants for change:

> Price's boom "Science since Babylon" has furnished some potent ammunition; the possibility that science may collapse of its own weight, and will do so much sooner with the thrombosis of its internal communicaiton is really beginning to worry some other people besides us![3]

With this remark, Lederberg introduced Price to Garfield. The latter was enthusiastic about the book: "I have now read Price's book, dutifully and enjoyably. I don't understand several statements he makes including some unsubstantiated remarks about electronic searching (though he may be right) but in general I think this is a wonderful study. I'd like to know that man better! I would be interested in his comments on the Citation Index and will probably write him for comment"[4]. Lederberg was advocating a centralized information system, modelled on John Desmond Bernal's plea from 1948[5]. He thought that in the end this might have resulted in the abolition of the traditional journals: "At first, the repository could coexist with conventional publication of the same titles, but it is obvious that the usual journals would wither away in competition with any really efficient service[6]. Lederberg also took this political development as an opportunity to promote the *SCI* itself: "The politics of this thing are too murky to let anyone see what possible constructive outcome may follow the committee's work; at the very least, though, I am sure there will be very strong support now for SCI; some of our friends are beginning to understand what it means and can do"[7]. It was no coincidence that Lederberg proposed strengthening his relations with ISI: "I

[2]Lederberg to Garfield, October 8, 1961. PSAC stands for Presidential Scientific Advisory Committee, better known as the Weinberg committee after its chairman, Alvin Weinberg.

[3]Lederberg to Garfield, October 8, 1961.

[4]Garfield to Lederberg, November 10, 1961. This led to intense communication between Garfield and Price right after the building of the *SCI* which laid the foundation of present-day scientometrics (chapter 4).

[5]"In my own mind, I am thinking of something like the ASTIA system of a repository, with broadcasting of abstracts and titles; we could then graft a much more penetrating retrieval system on to the repository. (Lederberg to Garfield, October 8, 1961).

[6]Lederberg to Garfield, October 8, 1961.

[7]Later, he asked Garfield to send 20 copies of his 1955 *Science* article to the secretary of the science information panel of PSAC (Lederberg to Garfield, February 9, 1062).

would now be delighted to accept your invitation to join the Board of ISI or NDC at your convenience"[8]. Lederberg asked Garfield to be his "informal consultant" and give him background information on "detailed proposals that you consider reasonably intelligent". Garfield showed no hesitation: "you can certainly count on me to be your informal consultant which means that I will probably spend a hell of a lot of time on what we classify as non-productive work, but I think the same can be said for you"[9]. At the same time, he cautioned Lederberg to be as precise as possible about the nature of the problems he himself had been thinking about for such a long time:

> For a long time I have been worried about the arrival of this moment—when someone like you would pop the question: well I'm listening so tell me what you think is wrong and what the solution is. In some of my own fantasies I wondered about this when I heard of some of the idiots who were appointed to such committees as you mentioned before. (...) I have been close to the SciInf problem now for a long time and this is what I find myself saying. After ten years I hear so little which I didn't know before. Like Bar-Hillel (a philosopher from Israel who pontificates much on Info Ret, etc.) I find myself saying "is there really a literature crisis" or is this something like a Stock-Market Crash Psychology which can plunge a country into economic disaster because it has lost faith in the market, etc. I think this is an important question that must be properly posed and answered. Is there really a crisis and if so what is the nature of the problem? I hear many people talk about Information Retrieval (IR) as though it were the burning issue of the century and yet I know few people who themselves really feel this to be true in their work.[10]

His conclusion was that the problem was different for different sorts of users: "Maybe the active producers feel the problem less than the consumers do". A few months later, Garfield reiterated his critical appraisal of the concept of information crisis: "Do you really know of any scientist who doesn't have enough time to read all the really pertinent papers in his field?"[11]. He thought that, notwithstanding the piling up of books, someone who did original research could still "keep ahead of the game". The real problem, Garfield felt, was finding the relevant documents, which very often took more time than reading the materials and thinking about their meaning.

The first thing the Weinberg committee set out to do was "go over the wording" of the report of an earlier committee, the Baker-report (PSAC 1958). This report had been the result of a national crisis triggered by the successful launch

[8]This should not be read as being inspired by the wish to steer possible projects to ISI. On the contrary, Lederberg acknowledged that his direct ties with ISI might make this "more difficult". His main motive was the promotion of the *SCI*: "You can see how upsetting this current experience has been to me, and I do not want to refuse any help I (or perhaps even my name) might give to some really constructive work such as you have been doing" (Lederberg to Garfield, October 8, 1961).

[9]Garfield to Lederberg, October 11, 1961.

[10]Garfield to Lederberg, October 11, 1961.

[11]Garfield to Lederberg, March 6, 1962.

of the Sputnik satellite by the Russians in 1958. Suddenly, higher education and scientific information were top political priorities. Until 1958, America's distrust of intellectuals had prevailed (Hofstadter 1962). The problem of accessibility of scientific information was not so much a broad concern of the scientific community, but had been confined to the world of the librarian. So, in 1949, the director of the Welch Medical Library at Johns Hopkins University called attention to the problem of "bibliographic control": "It is becoming increasingly difficult for our indexes and abstract journals to keep up with the growing number of medical publications and with articles of medical importance in other scientific journals" (Larkey 1949). The problem was at that time not yet recognized by scientific advisors of the President nor by those involved in organizing the new National Science Foundation. The Sputnik crisis turned the librarians' problem of bibliographic control into a national information crisis. The Baker committee was installed to propose solutions. It called for the establishment of a Science Information Service in the National Science Foundation (PSAC 1958, England 1982). It paid special attention to the way the Russians had organized their scientific system. The need to translate the vast amount of Russian literature was coined as one of the central problems in science information. Eventually, the panel decided not to adopt the idea of an all-American organization comparable to the Russian VINITI (the All-Union Institute of Scientific Information), because such a centralized institution did not fit in with the decentralized American system. Despite the Russian threat, the PSAC initially only proposed limited action. Major changes in the way science worked were not deemed necessary. Apart from the creation of a new division within the NSF, the panel made only one proposal: to investigate the application of machine methods and techniques.

> In other words, science must look within itself for a new system that will meet present-day requirements for the location, storage, and retrieval of scientific information.

This would be a constant theme in the American solution to the information problem. The Baker Report was nevertheless quickly seen to be utterly ineffective. As Lederberg told Garfield, the Baker committee "fathered that absurd report that was issued 3 years ago, encouraging a coordinative role for the NSF, and anarchical free enterprise"[12].

The new PSAC "Panel on Science Information", of which Lederberg was a prominent member[13], issued its report in 1963 with an array of proposals and calls for action, directed at the federal government, the scientific community, in-

[12] Lederberg to Garfield, October 8, 1961.

[13] Alvin Weinberg, director of Oak Ridge National Laboratory was the panel chairman. The other members were: William Baker, who had chaired the previous PSAC panel on information, Bell Telephone Laboratories; Karl Cohen, General Electric Company; James Crawford jr., editor *Journal of Applied Physics*; Louis Hammett, Columbia University; A. Kalitinsky, General Dynamics/Astronautics; Gilbert King, IBM Research Center; William Knox, Esso Research & Engineering Company; Milton Lee, Federation of American Societies for Experimental Biology; John Tukey, Princeton University and Bell Telephone Laboratories; Eugene Wigner, Princeton University; Jay Kelly, Office of Science and Technology, Executive Office of the President (PSAC 1963, 51).

dividual scientists and the libraries[14] This report differed from that of 1958 in that it found that the information crisis was not merely a question of keeping the individual scientist informed. The crisis was threatening the very identity of science. The report opened with the following sweeping statement:

> Science and technology can flourish only if each scientist interacts with his colleagues and his predecessors, and only if every branch of science interacts with other branches of science; in this sense science must remain unified if it is to remain effective. The ideas and data that are the substance of science and technology are embodied in the literature; only if the literature remains a unity can science itself be unified and viable. Yet, because of the tremendous growth of the literature, there is danger of science fragmenting into a mass of repetitious findings, or worse, into conflicting specialties that are not recognized as being mutually inconsistent. This is the essence of the "crisis" in scientific and technical information. (PSAC 1963, 7)

So, in the course of twenty years the nature of the information crisis had changed from a librarians' problem of bibliographic control, to a problem for the individual scientist trying to cope with the growing volume of literature, into an identity crisis of science in general. At the same time, as President John F. Kennedy wrote in his foreword to the report, science itself had become "a national necessity" (PSAC 1963, III). The Weinberg report called for drastic action and for major changes within the scientific system.

The Weinberg Committee

One of the committee's recommendations was the development of a new searching tool — the citation index — about which the panel was "particularly impressed". The traditional ways of making literature available to the scientist were collapsing, revolutionary changes were necessary.

These recommendations were at least partly the result of an intense correspondence between Lederberg and Garfield about the solution to the problem of scientific information while they were constructing the *SCI*. Garfield had immediately devoted himself to the problem: "During the last month I have been trying to separate myself from what I am doing on a day-to-day basis in order to arrive at some conclusions that might be useful to you in your work on the Advisory Committee"[15]. Again, as he had done in his previous discussions about citation indexing with William Adair (see section 2.2.2 (page 23)), he put computers central stage.:

> I was prompted to think out the cost of putting into machine language, i.e. on magnetic tape or some other medium, the main scientific output of the world. If you assume that the average scientific article is about 5,000 words and a girl can consistently type 25 words per minute of scientific text then the cost of putting an article in machine language is about 3 man-hours or

[14] PSAC (1963) has since been seen as a landmark in the history of documentation (Schneiders 1982, 176).

[15] Garfield to Lederberg, November 10, 1961.

> at going rates about \$5. Add a safety factor and other costs and maybe this would be as high as \$10. If there are 1,000,000 useful scientific articles per year then this is \$10,000,000. Among other by-products of this very simple operation we would find the following possibilities:
>
> a Permuted title indexes
>
> b Citation Indexes
>
> c Mechanical translation of foreign texts
>
> d Mechanical analysis of texts for indexing and retrieval
>
> e Miscellanous other by-products

Garfield recognized that mechanical translation was "still somewhat primitive", while he also thought it would take time before mechanical textual analysis would be practical. Nevertheless: "regardless of the imperfections of all of these I think that we can expect that high speed computers will be able to process this material at a cost that is not higher than the input cost of \$10,000,000 and probably is lower". This would be reflected in the creation of new "departments of science information" at every major university ("preferably not tied to the library school"), as well as a solid training in science history[16].

Garfield laid out a comprehensive scheme to Lederberg comprising "three levels of reporting": "Title, Abstract, Full paper". The basic idea was that one per cent of the papers would be published in a "national or international organ" (for example a daily science newspaper), the next ten to twenty-five per cent would be published in "a series of select journals". The vast majority of the papers would be put in a central depository. This would put an end to the proliferation of new journal titles. The newspaper would also publish lists of all papers: "A national organ appearing five times per week could list the titles of one million papers at the rate of 4,000 per day on about 20 NY Times size pages"[17]. The national documentation centre, which would be the central axis, would distribute "a series of abstract journals". Moreover, a "prompt translation service" would provide for fast international communication. Garfield envisioned his *Current Contents* or its successor as *the* place to "publish by title", whereas the newspaper would also have a daily citation index section. All in all, the system would be a drastic improvement for timely access to all available information. "An important factor is", Garfield stressed, "that a man's personal bibliography should have the same publication value regardless. A reference to a paper that does not get into the major primary organ or in the journals should be considered equally"[18]. Garfield reiterated Bernal's ideas, as he made clear by urging Lederberg to look into Bernal's papers[19].

Timeliness "in everything we do" was the key to Garfield's vision of the bibliographic future: "if we are going to have information services—let them really

---

[16] "I have always felt that we needed greater stress on history of science and my interest in Citation Indexes derived in part from this." (Garfield to Lederberg, November 10, 1961).

[17] Garfield to Lederberg, November 10, 1961.

[18] Garfield to Lederberg, November 10, 1961.

[19] Garfield to Lederberg, November 15, 1961.

be prompt—let them really be comprehensive and consistent—and let there be immediate access to documents"[20]. In the end, Garfield thought, there would be only two kinds of scientists, "information scientists" and "the lab men", assisted by computers "to really generate new information from what we already have".

Bernal's stamp was also clearly visible in the first draft Lederberg wrote of his "Notes on a Technical Information System" (Lederberg 1962*a*)[21], reconstructing the main problem as follows:

> As members of the scientific community we have a deeply rooted obligation to interact with the "literature". Not so much the size but the dispersion and formlessness of the institution make this an ever more hopeless aspiration. (...) The present system has generated two responses: the defeat of neurotic frustration for some, the compromise of narrow specialism for others. I feel the survivorship of humanistic science demands a better solution. (Lederberg 1962*a*, 1)

Lederberg proposed devising a central repository in combination with "select journals":

> A centralized repository would provide the range of materials that I would specify as being required for my immediate and retrospective information requirements. Concurrently, select journals with high standards of selection and editorial quality would maintain my contact with the breadth of scientific culture. (Lederberg 1962*a*, 7)

The repository would be built according to a set of ground rules. One of these would be that no paper could be withdrawn once deposited "as with journal publications the author's reputation is permanently attached to it"[22]. Papers would be distributed and refereed "promptly". Moreover, an updated citation index would be attached to the articles. Garfield liked this idea:

> Your idea of an updated citation index going out with each copy of an article is the greatest!! I had overlooked this very obviously good service. Fabulous it is so simple.[23]

The principal advantage of the repository scheme was, according to Lederberg, the "prompt and widespread availability" of contemporary findings. "That contributions can take a full year to come out in print is an absurdity of modern science" (Lederberg 1962*a*, 4). The repository would "discourage the redundancy

[20] Garfield to Lederberg, November 15, 1961.

[21] In this draft, Lederberg clarified his ideas by applying his schedule to the NIH community and his personal information needs as a geneticist.

[22] Lederberg specifically mentioned the advantages of adding notes to the publication by the author: "The author, of course, may submit amendations, corrections, etc., to be attached to a previous submission. The possibility of doing this is already a substantial advantage over present publication means" (Lederberg 1962*a*, 3). In recent decades, virtually the same set of rules has been proposed in relation to electronic publishing (Harnad 1991, Harnad 1990, Odlyzko 1995).

[23] Garfield to Lederberg, February 20, 1962.

implicit in peripheral publication and in the irresponsibility of gossip and 'invisible colleges"'. It would facilitate the publication of "expensive archival documents" like taxonomies. Last but not least, it would stimulate the journals to "revert to being select journals": "they are broadsides on which I would rely to bring me unasked the best or overtly most interesting of contemporary science". The user would be more central than in the prevailing system, Lederberg felt. He expected the journal output to decrease to about 10 per cent of its current level.

The central problem in realizing this radical overhaul was that it needed a certain critical mass, Lederberg wrote to Garfield:

> I have been very much worried how this could be brought to its necessary critical mass to prove itself since many authors may be reluctant to be the first ones to use the depository and distribution on call as an alternative to journal publication. The newspaper distribution would, I believe, rapidly build up interest in the system at which point it should gradually become converted into a vehicle for the publication of titles and indices and news of science rather than the typical articles themselves.[24]

After having discussed these items with the NIH's Mental Retardation Panel[25], he became even more convinced of the necessity of a newspaper for science.

Hence, the daily newspaper they had discussed earlier (see section 2.3.7 (page 47)) was not only the crucial connection between the *SCI* and the centralized information system, it also became a strategic item in realizing their information revolution. The publication system already in existence, with all of its vested interests, seemed to be the main obstacle:

> I wouldn't necessarily want to lick the profit making journals, but I think they should get out of the business of primary documentation. They will still have plenty of work to do in organizing review journals and quite possibly some profit incentives may be important or necessary in getting people to do the work of editing and distributing to journals of this kind. But these should have the status of secondary compendia, or books[26]

"Unfortunately", Lederberg noted, "one of the serious shortcomings of the OSIS[27] in NSF is that it really has neither the staff nor the mandate to consider such large scale systems propositions". Garfield had the same experience. He had sent his proposal for a unified index to science in newspaper format to NSF but had not heard since[28]. He sent Lederberg some computations about the needed critical mass for a newspaper for science to be successful. He also proposed a tactical opening to break the deadlock of the prevailing system:

---

[24] Lederberg to Garfield, February 19, 1962.

[25] Lederberg to Garfield, February 26, 1962.

[26] Lederberg to Garfield, February 26, 1962.

[27] The Office of Scientific Information Service.

[28] "This proposal was sent quite informally to Alice Billingly who is on the staff of NSF Office of Scientific Information Service. Like a good litlle bureaucrat, she has done nothing about it — not even to write me a note suggesting whether this is a good time to formally submit it." (Garfield to Lederberg, March 6, 1962).

> If we assume that the members of an organization, such as the National Academy of Sciences, consist of people who are agreed that the present communication system is not good, and that these people do not have as much of a need for this publication as do junior scientists, and these same men would like the idea of reaching a mass audience of scientists and laymen, isn't it possible that with their participation, or a majority participation, the junior members of the scientific community would then quickly jump aboard?[29]

Garfield[30] wrote of the possibility that the Nobel Prize winners would sign some form of pledge in which they would agree that they would publish all their original research in the new, condensed form in the National Newspaper of Science[31]. He seemed a bit more optimistic concerning the resistance of commercial publishers: "if we do stress that future developments will only accelerate what is already taking place, then this entire new approach can be extremely profitable to publishers". In Garfield's opinion this was particularly true because of the rate at which science was expanding. "In other words, if our proposals for such enterprises as the Daily Scientist would cut down on the amount of journal publication, it might achieve this effect by at least only preventing the birth of new journals. So-to-speak, this is a kind of journal birth control and gradually other older journals would die out."[32] In other words, Garfield was convinced that the market for information services was expanding.

The SCITEL System

June 1962, Lederberg wrote a memorandum to the Science Information Panel of the Weinberg committee in which he unfolded his ideas on "a central scientific communication system" (Lederberg 1962*b*). It was an unequivocal plea for a centralized system: "Responsible submission of fully documented papers to a central repository is the primary act of scientific communication" (Lederberg 1962*b*, 1). The effective pursuit of science called with growing urgency for "efficient, systematic, anxiety-free, reliable access to the exponentially increasing flow of scientific information". The repository would both function as an archive and as a switching center, authors would send their articles to the repository instead of submitting them to journals. "Deep retrieval services" would enable easy access to them and were the key to the integrity of the system. Broadcast services would alert scientists about new materials, consisting of current announcements by title of receipts to the repository, cumulative indexes and a daily journal. The daily journal had a strategic role in Lederberg's scheme. He thought that it would be "the main leverage to accelerate the SCITEL system" and would have the form of an expanded analogue of the *Proceedings of the National Academy of Sciences*. It would become "the principal conventional vehicle of primary dissemination of

[29] Garfield to Lederberg, March 6, 1962.

[30] Garfield felt that he and Lederberg operated on "exactly the same frequency".

[31] "The full text of their research would, of course, be available through the depository." (Garfield to Lederberg, March 6, 1962).

[32] Garfield to Lederberg, March 6, 1962.

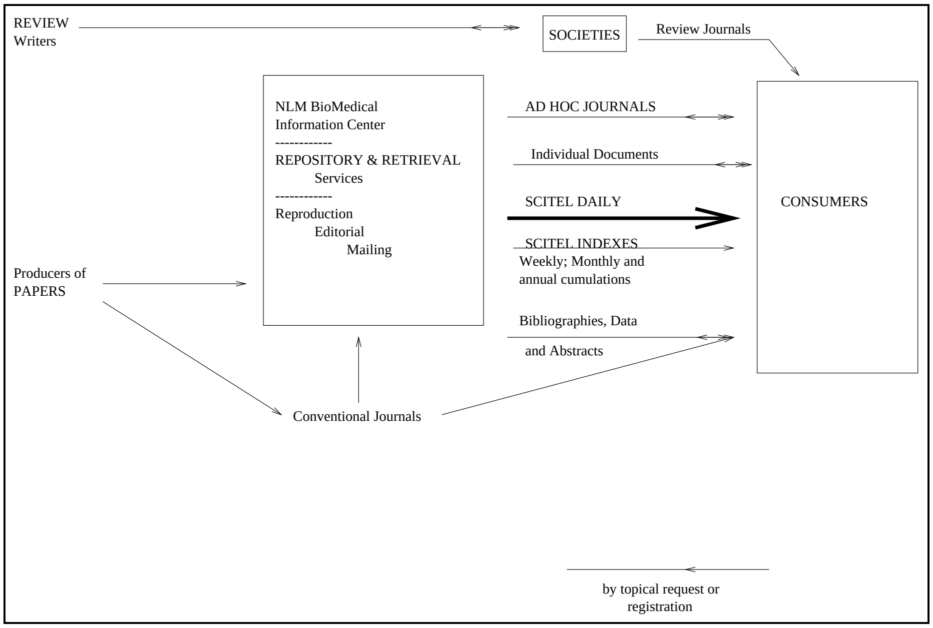

Figure 3.1: Lederberg's SCITEL proposal.

documents from the center". Moreover, it would be the "habitual locus" for indexing material and news of science. In short, it would represent a "sample of current science" (Lederberg 1962*b*, 2). Whereas the journal was expected to publish only the top publications[33], all documents would, once deposited, receive an acquisition number by which it could be identified. It could then be distributed on request as a separate reprint.

Lederberg's scheme (figure 3.1 on page 68) differed fundamentally from conventional publication in scientific journals. First of all, the primary responsibility for seeking editorial criticism would be shifted to the author. Second, the need for primary journals would disappear. "Relieved of the unnatural responsibility for primary archives and communication", the scientific societies and other journal sponsors could devote themselves to too often neglected services "especially in review and interpretation". At this stage of scientific communication, Lederberg wished for more commercial opportunities:

> In this area authors and editors may deserve royalties for incentive and reward and the free enterprise arguments are generally most valid. But there will now be a fair market for subscribers' choices without the coercion now implicit in the need for access to the primary literature. So there will not be the inordinate pressure for scattering and increase in journals, only what the market wants and can actually afford. (Lederberg 1962*b*, 4)

Third, authors would also be responsible for the production of abstracts, since manpower requirements prevented their central production. Lederberg acknowledged the possibility that "peripheral agencies" might also be able to continue

[33]Lederberg estimated that between ten to thirty per cent of the papers in a field might be published in the Daily Scientist.

their abstracting services. Fourth, the government would have the primary responsibility for financing the whole system. Lederberg was optimistic about these economic aspects:

> The actual cost of a centralized mechanism must be considerably lower than the present journal system counting only the savings in efforts of editors and librarians and the economics of large volume printing. Its impact on scientific efficiency should be the main consideration. These savings may already counterbalance the wastage involved in sending more newsprint to every scientist he can likely read in detail.

Fifth, the system would be oriented to innovation, looking to "the future development of data handling and telecommunication systems to replace the techniques of the present proposal".

ISI's experiences with *Current Contents* was a great help in getting a grip on the citation index project, it also led Garfield to increase the scope of his operations. This was not only the consequence of the promising technical results. Garfield's deliberations with Joshua Lederberg also stimulated thinking in terms of shaping the future of scientific communication:

> I have been thinking "big" down here in terms of ISI's future. I hope to incorporate this thinking into a series of proposals that tie in with your proposals on Science Advisory Committee. (...) I am convinced we are only five to ten years away from bridging the existing artificial gap between technical science writing and writing for the laymen. In fact, there is probably a greater need than you and I realize for a citation index "structure" that would relate a conventional clipping service with our scientific clipping service.[34]

This included a possible central role for ISI: "Have you and enough like you reached the point where "printed publication" is no longer important? If a dozen men like you began to "publish" via CC-title and ISI depository, how long would it take for others to follow suite?"[35] While the computer programs, data files and citation indexing procedures were developed, the question of publication the resulting index became more pertinent. Garfield proposed NSF test "the newspaper format" for a daily citation index[36], in order to achieve a "low cost per reading". The newspaper should have the format of *The New York Times*, initially comprise sixteen pages and bring reprints of original research papers and review articles (four pages), a daily updated author bibliography (five pages), a citation index (six pages) and a subject index (one page) (Garfield 1962*a*). The author bibliography would contain 750 papers per day and was vital for the use of the indexes. Garfield expected that in one year three million citations would have been listed this way. The "Daily Scientist" as it was called should be a throw-away paper: "The philosophy behind a daily dissemination technique is that the information comes in small segments. The daily newspaper is quickly scanned and then discarded" (Garfield 1962*a*, 2). Garfield estimated that scientists would be

[34] Garfield to Lederberg, July 9, 1962.
[35] Garfield to Lederberg, July 9, 1962.
[36] Garfield to O'dette, September 17, 1962.

prepared to pay a subscription fee of thirty dollars a year. He proposed NSF test the idea by sending 25,000 scientists consecutive daily issues for two months. If NSF would give initial support, the experiment could be expanded with the help of NIH, NASA and AEC. Garfield estimated that a one-year experiment would cost around 500,000 dollars. Most of this money would be necessary to produce a Unified Citation Index to Science anyway. Therefore, Garfield could maintain his claim that it would "bring a vast amount of information to the individual scientist at a phenomally low cost" (Garfield 1962*a*, 2).

3.1.2 Technical design

Building the *Genetics Citation Index* was a huge operation. Not surprisingly, it was plagued by delays and budget problems. It started slowly. 1961 was a year full of other obligations and challenges for Eugene Garfield: "Every day brings exciting new ideas"[37]. For example, ISI was in the race for a three million dollar contract with NASA. It made Lederberg worry about the citation project, while Garfield saw it as an opportunity to extend the journal coverage at ISI to physics journals[38]. He assured Lederberg that he was not stalling: "I am trying to get up real movement on the Citation Index Project but other distractions have prevented it"[39]. Because of Garfield's managerial and other obligations, the advisory committee did not meet before December 1961. External circumstances also took their toll. Garfield had to be the actual project supervisor for a year, because of the departure of Gwenn Bedford[40]. Only in February 1962 could Garfield inform Lederberg that he had finally solved the project management problem: "The big news is that we now have a full-time project director—Dr. Irv Sher. A smart fellow you will want to meet some time. He was at SK&F[41] and he will help me a great deal. The project is now much too big for me to handle alone"[42]. The contract with NSF was not completed before May 1961 and as a consequence, the punched card equipment was coming in only slowly. Garfield and O'Connor concentrated on investigating "in considerable detail" the various ways of compiling the index[43]. Garfield felt that the first year should be devoted to systems studies, especially since the NSF funding allowed the desired comprehensive approach[44]. As he stressed in the first informal Progress Report: "The Citation Index concept is beautiful in its basic simplicity. However, its own peculiar complexities and ramifications require careful analysis" (Garfield 1961, 1–2). These complexities related to important and unsolved issues of technical design decisions, as well as to unknown parameters of the prevailing citing cultures in scientific communica-

[37]Garfield to Lederberg, July 5, 1961.

[38]Garfield to Lederberg, June 19, 1961.

[39]Garfield to Lederberg, July 5, 1961.

[40]He hired part-time assistance in the person of John O'Connor for one day a week, and was also assisted by Catherine Voytko, but this did not relieve him of his supervising responsibilities.

[41]Smith, Kline & French, a pharmaceutical company Garfield had been consultant for.

[42]Garfield to Lederberg, February 17, 1962.

[43]Garfield to Stewart, July 26, 1961.

[44]"In January, I decided that we should devote as much time as possible during our first year to system studies." (Garfield 1961, 1)

tion.

Bibliographic discoveries

Both Garfield and Lederberg were very interested in information theory. Garfield combined this with an active interest in linguistics and coding problems in general[45]. He tapped Lederberg's knowledge of these topics, especially in relation to the genetic code[46]. He also made some independent discoveries. Garfield was the first to describe the possibility of co-citation analysis, although with a different aim to that of Henry Small and Irina Marshakova who developed their technique in the 1980s:

> I figured out a method one can use to decrease the noise of a citation index search. If you find that a particular reference has too many referants[47] associated with it then you can think of another reference which also may have a long list of referants. Then you match the two sets of referants and this gives you a more specific search. Put another way—what articles are there that have referred to a particular pair of references. If I specify that a reference shall have been made both to a paper in genetics and to a particular paper in information theory the referants may give me just those papers concerned with the problem of information theory applied to genetics!! This can be done very easily on a computer and corresponds to a widely used method known as coordinate indexing.[48]

He did not publish this however[49].

Garfield discovered how references and citations were distributed in the literature which sometimes made him anxious about the scale of indexing literature. After Lederberg had asked Garfield to list citations to Lowry's and Kety's work, he became quite worried: "He really hit the jackpot with his article in the JBC, 193, 265, 1951. I have a sort of panic about this sample and wonder whether this can be useful to anyone"[50]. By that time (beginning of 1962) Garfield had seen 6,000 pages of printout of the *SCI*, which "absolutely convinced" him that the index could not fail as a current awareness service. Unless, and this is why he was so worried about the excessive number of references to Lowry's article, the indexing project were to get "logjammed by excess volume":

> This sample has given me an opportunity to see first hand the effect of large quantity. It is just a huge mass of printout—about 8000 pages. Even the

[45]Garfield had written his Ph.D. thesis on this subject.

[46]Garfield to Lederberg, July 5, 1961; Lederberg to Garfield, July 26, 1961.

[47]A "referant" is the citing document; a reference is the cited document.

[48]Garfield to Lederberg, July 5, 1961.

[49]This partly reflects the practical orientation Garfield had to these problems. Perhaps the not very enthusiastic reaction of Lederberg to this particular idea also played its role: "Is "coordinate indexing" a new idea in SCI? But it doesn't eliminate "noise" which usually implies error." (Lederberg to Garfield, July 26, 1961).

[50]Garfield to Lederberg, February 6, 1962 (the first page of this letter reads "1961", clearly a typo).

> disposition of the sample printouts is a physical problem. I've got the sheets all over my office.[51]

Lederberg apologized to Garfield: "I should have briefed you about the Lowry paper. It probably is the most frequently quoted article in biochemistry since it refers to a now standard method for the analysis of proteins. Actually it is a little silly for it to be so constantly quoted since "the Lowry method" would be quite sufficiently indicative even without the reference"[52].

The sheer amount of work turned out to be the main source of delay. For example, the original proposal had been a little too optimistic about the number of journals which could be processed. At the end of 1962, of the projected one thousand journals only 440 had been processed[53]. All in all, one hundred thousand cards had by then been punched.

Garfield discovered also that the *SCI* could form the basis of a personalized clipping service for scientists:

> Is it reasonable to assume that if I cite a paper that I would probably be interested in those paper which subsequently cite <u>it</u> as well as my own paper. Indeed, I have observed on several occasions that people preferred to cite the articles I had cited rather than cite me!! It would seem to me that this is the basis for the building up of the "logical network" for the citation index service.[54]

Publishing commercially

Compiling the index proved to be more expensive than expected. The second progress report (Garfield 1962*b*) confirmed that "large scale citation indexing is practical"[55]. At that moment, about 1.4 million cards had been punched "of which 1.25 million represent usable reference citations"[56]. Garfield reported proudly that he had been able to reduce the indexing costs to $3\frac{1}{2}$ cents per card. On the other hand, processing foreign journals and especially standardizing "thousands of journal abbreviations" was much more costly than foreseen. The project was spending $ 15,000 a month. For this reason, Garfield requested an amendment of the contract between ISI and NSF "to include at least $ 90,000 for six months" and preferably $ 270,000 in order to complete the full research program. It was the start of a period of bickering about money. Garfield also asked for $ 100,000 to print the index, amounting to a total budget for the year 1963 of $ 380,000[57]. He was not very optimistic about the future: "It is obviously the intention of these people to cut us off at the earliest opportunity, though I do not think this is necessarily based on malice". NSF did display interest in the Daily Scientist idea

[51] Garfield to Lederberg, February 6, 1962.

[52] Lederberg to Garfield, February 9, 1962; Lederberg to Garfield, February 19, 1962.

[53] Sher to Rhodes, Adkinson, Sprow, Garfield, Riesenbach, Schiller, Anzlowar, Heatwole, November 20, 1962.

[54] Garfield to Lederberg, February 2, 1962.

[55] Garfield to Sprow, September 25, 1962.

[56] Garfield to Sprow, September 25, 1962.

[57] Garfield to Lederberg, December 20, 1962.

as a way of pursuing the citation index project. However, it clearly did not wish to finance the production of a citation index on a regular basis. This steered the citation index in a commercial direction. Garfield started to seriously consider publishing the *SCI* of his own accord:

> Since NSF has frustrated any plans we had for even limited distribution of the entire 1961 file, and since they have, through Ralph O'Dette, indicated that there would be no objection whatsoever to our publishing this information, we are now considering in detail advantages and disadvantages of publishing this file as soon as feasible. Depending upon the format selected, it will probably be an index of about 5000 pages. We propose to complete the coverage of the life science journals with the NSF funds and then to complete selected physical science journals with our own funds.[58]

Garfield expected a price in the range of $ 1000 and $ 5000, the profits would be re-invested to produce the citation index for subsequent years. Worries about the competition speeded up these decisions: "We are quite concerned that we may have already advertised and exorcized the virtues of citation indexing to the point where potential competitors who are better financed than ISI can jump in ahead of us"[59]. Garfield hoped to find a partner who could put up "from 1 to 2 million dollars" and was patient enough to realize that it would take "from 3 to 5 years" for a science information service to become profitable. In the end, the *SCI* was mostly financed by the profits *Current Contents* generated.

3.2 Translating the citation concept

3.2.1 Automation

The citation index NIH and NSF supported and the *SCI* as it would be published from 1964 onwards, did not look like Shepard's Citations anymore. Technically, the idea was still the same. Because of this, Garfield's proposal to NSF could state that most of the uses of the *SCI* were "analogous to their use in legal research". This statement nevertheless concealed essential dissimilarities. The fundamental change was in the meaning of a citing relationship. The outlook of the index differed as well. Moreover, ISI's way of producing the index would be the complete opposite of Shepard's. The production of the *SCI* was therefore not a matter of simply applying a ready made tool in a new area. Developing the *Science Citation Index* required both a new way of looking at scientific literature, and a new conception of citation indexing.

Shepard's refusal to expand its market beyond the law libraries, was not a matter of shortsightedness. Producing science citation indexes would have meant overhauling the whole firm, thus jeopardizing Shepard's profitable serving of America's legal system. In deciding not to do this, the management showed its sound business instinct. In the fifties, computers were developing quickly,

[58]Garfield to Lederberg, December 20, 1962.
[59]Garfield to Lederberg, December 20, 1962.

but they were still unwieldy and difficult to handle. Only the enthusiastic to the point of devotion saw the point of manipulating vast numbers of punched cards. Garfield was such a person, keen on looking for computerizing all sorts of processes. As it happened, the mechanical production of the citation index without the need of experts being present was decisive in the advent of this new concept in the realm of scientific publication. Automation was, however, only made possible by stripping the citation of every judgment on its meaning.

> The brilliant utility of the citation index approach is that it cuts across the problem of meaning by an automated procedure.[60]

Shepard's classified its citations with an array of symbols; for example, whereas an *a* ment an affirmation of the decision, an *r* indicated that the case was reversed. Initially, Garfield tried to develop a comparable meaningful classification scheme for science. But Joshua Lederberg immediately made it clear that this would be self-defeating. Not only because abstracting services already existed, but also since these services had years of backlogs. After all, they had to make judgments on the scientific literature they processed. This relates to an important distinction between dealing with scientific and legal literature, i.e. the magnitude of the enterprise. Scientists simply produced much more papers. Moreover, as Gordon Allen pointed out, science as a whole was in a far from steady state and seemed to be growing exponentially.

3.2.2 Comprehensiveness

When he was discussing citation indexing with Garfield (chapter 2), William Adair had also been aware of the problems of scale. He had not considered automation, however. Instead, he had proposed the separate indexing of the various scientific specialties or disciplines. It was a familiar solution; Shepard's Citation was also classified according to the structure of the legal system in the US. Neither was it a new idea in the world of science. After all, most scientific journals were limited to a narrowly defined specialism. Moreover, several other citation index projects were constructed along the same lines. Garfield had been the principal propagandist of citation indexing[61] but he was not the only one involved by any means. In the early 1960s, NSF supported several citation index research projects. The *SCI* project was, however, the only attempt to produce a comprehensive citation index covering, in principle, all of science. The two principal "competitors" had chosen explicitly in favour of a monodisciplinary approach. Statistician and leading citation index researcher John Tukey studied and built a citation index of the statistics literature at Princeton University (Tukey n.d.*b*, Tukey 1962, Tukey n.d.*a*). This choice was made because

> to cover all that interests workers in a field is a surprisingly broad task, as the experience of the ISI-genetics project, which has roughly followed this

[60] Lederberg to Garfield, November 9, 1962.

[61] As John Tukey wrote in Tukey (n.d.*b*) "A large part of the recent attention to citation indices has come from Eugene Garfield."

> road for the field of genetics, shows so clearly. In the case of statistics, such a decision would bring in not only all of mathematics, but large and rather ill-defined portions of most scientific disciplines and many technological specialties. (Tukey n.d.*a*, 29)

At the Massachusetts Institute of Technology, the inventor of the concept of bibliographic coupling, Michael Kessler (Kessler 1961), was constructing a complete information system of physics literature. He did not consider a citation index strong enough to sustain a pilot model system in itself, though it would be a useful element to add once the model was constructed, because citation was "a low probability event" (Kessler & Heart 1962, Kessler 1965)[62].

By automating the production of the *SCI*, Garfield, Lederberg and Allen could tackle the enormous task of indexing the scientific literature while retaining its complete coverage of science. The objective of the *SCI* to cover all of scientific literature was underpinned by the concept of "the unity of science". Without the possibility of going beyond the boundaries of the academic disciplines, a citation index would add practically nothing to the traditional subject indexes. After all, researchers could be relied upon to be familiar with the literature of their own specialty. The merit of the *SCI* was its ability to locate relevant research in unexpected places. This seemed only possible if the *SCI* were not structured along rigid discipline-bound lines. The *SCI* was also expected to change the citing behaviour of the scientist. This was not the case with Shepard's Citator as it was only a registering device. The citing behaviour of attorneys and judges was fairly standardized. This made it possible for the indexers to classify the citations using a fairly restricted set of symbols. In contrast, scientists are far less restricted in their citing behaviour. References to scientific papers play differing roles. Even the same citation can change its meaning in the course of time, the meaning for the author not having to be the same as the meaning to the reader. The makers of the SCI expected to exert a positive influence on the scientists' citing behavior. In its turn, this would increase the value of the *SCI*. Lederberg expressed it as follows:

> You can be sure that if you set up CI for citations to Science, Nature etc. that many authors will then take care to include more references to these journals which will help to ensure better coverage of the literature.

3.2.3 The information crisis

In the fifties, science had increasingly been felt to be growing too fast to be able to cope with its results. It had made some parts of the scientific community gradually more receptive to innovations in handling the literature. This "information crisis" is a key factor in the birth of the *Science Citation Index*, playing social as well

[62]Kessler & Heart (1962) was vehemently disputed by Garfield, Lederberg and Tukey (Lederberg to Kessler, January 21, 1963; Kessler to Lederberg, January 28, 1963; Adkinson to Lederberg, January 29, 1963; Kessler to Garfield, January 28, 1963; Garfield to Kessler, February 27, 1963; Tukey to Kessler, February 11, 1963; Garfield to Tukey, February 16, 1963).

as cognitive roles[63]. This information crisis shaped the way the central problems in the realms of science, science management and science policy were defined. Government agencies provided funds to find solutions to this information crisis and thereby created a new labour market, asking for people with both scientific and librarian skills. This new field was the province where people from such diverse backgrounds as a documentation specialist, a researcher in human genetics at the National Institutes of Health, a Nobelprize winner in bacterial genetics and a retired vice-president of Shepard's could meet each other. The crisis, made more urgent by the Sputnik crisis, finally gave citation indexing the offical approval it needed to take off.

As has been said before, a debate at NIH about the evaluation of NIH-funded research reminded Lederberg of Garfields 1955 paper in *Science* and prompted him to write his memo in 1959. Once a citation score is transformed into a measure of impact of a paper, if the database is large enough all sorts of policy related studies can be easily envisioned. The sociological use of the *SCI* was an outgrowth of this and of the network approach. Notwithstanding this, the central motive for scientists like Lederberg and Allen was, and would remain, the search capabilities the *SCI* gave the user. They were very suspicious of possible misuses of the *SCI* for science policy ends. Lederberg made this quite clear on the brink of *SCI*'s publication. In January 1963, he expressed his opposition in a letter to Irving Sher "to any proposals for the use of citation index statistics in personnel evaluation"[64]. He explained that he had two reasons. First, not enough was known about the citation structure in the literature:

> Ultimately, we may know enough about the structure of scientific communication to be able to use this kind of information intelligently, and especially to apply the necessary kind of correction factors needed for such a purpose, but until then, and I think this is a long way off, the idea of such a statistical evaluation is a dangerous one

Second, it might harm the citation index project itself:

> if the misunderstanding gets around that this is an implicit objective of the citation indexes, it is likely to arouse a great deal of hostility on the part of the scientific community, and this may not be always entirely rationally directed.

Lederberg did see a role for the *SCI* in personnel selection, but only because it would facilitate the location of commentaries or critical appraisals of a person's work[65].

[63]Of course, this crisis is a cultural product as well. It was not a natural conclusion to look to "information" as the domain where the solution was to be found. I do not question the structure and character of the "information crisis" at this level. Evidently, it is a consequence of the way the US drew conclusions from World War II — and the Manhattan Project — in Vannevar Bush' report *Science as an Endless Frontier*. Interestingly, Garfield himself was also rather critical of the notion of an information crisis.

[64]Lederberg to Sher, January 7, 1963.

[65]Lederberg essentially sticked to this position over the years (Joshua Lederberg, Personal Interview, February 3, 1992, New York).

3.2.4 Computers

Computers play a tremendously important role in this history. Without computers it would have been simply impossible to make a database such as the *SCI*, because it would have been far too expensive. Even with the existing computers it was a risky business. It was the computerized processing that made the migration of the citation concept from the legal to the science context possible. The corresponding devaluation of labour made the production of the *SCI* possible within the budget available for these kind of enterprises in the US at the time.

3.2.5 Innovative outsiders

Without the drive, perseverence and social capacities of Eugene Garfield as well as the technical expertise with which he surrounded himself, the immense task of building the *SCI* would probably not even have been thinkable. It is not only a matter of personal traits, but also of being at the right place at the right time. Garfield was an outsider in more than one respect, which enabled him to think about solutions other people would reject immediately. And he definitely had a taste for information services, *Current Contents* being the proof. Not coincidentally, the two scientists who reacted to Garfields 1955 article in *Science* were geneticists. The structure of the new science of genetics made coping with the literature more pressing for Lederberg and Allen than for, say, the nuclear physicists. Genetics still had undefined boundaries. The professional societies of human and bacterial genetics however stuck to the old subject indexing.

In the process of translating the citation concept to the world of science, the funding agencies and Eugene Garfield learned to get along with each other. Garfield was an outsider in the academic world[66], running his own company, Eugene Garfield Associates, with *Current Contents* as its main product. He was not affiliated to an academic institution. An intellectual problem existed as well. Citation indexing was unknown territory. Garfield's proposals showed this and, naturally, he wanted to keep as many options as possible open. The funding agencies were also uncertain and wanted to know more precisely what they were supposed to be funding. The support of Allen and Lederberg made Garfields undertaking more acceptable to the funding agencies.

3.2.6 Success as well as failure

The experimental genetics citation index appeared in 1963, the *SCI* in 1964. Since then, the *SCI* and its associated products have become a well-known feature of every scientific library in the world. ISI almost broke down because of the *SCI*[67], but in the end the index turned out to be profitable. It seems the classical American

[66] Which does not mean he did not have many contacts with researchers and science policy officials. On the contrary, networking is one of Garfield's strong points. He was moreover asked to review proposals to NSF on indexing projects on a regular basis (Garfield to Gray, September 12, 1959).

[67] Eugene Garfield, Interviews, January 27, 1992, and February 4, 1992, Philadelphia.

success story, log cabin (Garfield's chicken coop in New Jersey where he started producing *Current Contents*) and all. And a success story the *SCI* certainly is.

But it is also a story of failure. Lederberg was not only thinking about a bibliographic tool when he pushed the case for citation indexing through the courts of science policy. He set out to revolutionize the whole publication system of science. In 1959, Lederberg had adopted Bernal's scheme of doing away with all scientific journals as a primary channel of publication. As a member of the PSAC panel on scientific information, having been impressed by Price's book *Science since Babylon*, he pressed for the abolishment of the anarchist way of publishing. He thought that all commercial publishers should be pushed out of the business of primary publication. In association with Garfield, who was acting as his informal consultant on the matter, Lederberg developed his ideas along these innovative lines. In their hands the *SCI* would not be merely a searching tool, but a revolutionizing instrument, profoundly changing the world of science. In this respect their enterprise was a failure. The birth of the *SCI* did not bring about any immediate changes in the scientific community, nor did it profoundly influence scientists' behaviour. By limiting the scope of the *SCI*, the existing institutions succesfully defended the traditional way of publishing. The *SCI* became an important but somewhat peculiar product of a company with an interesting, but somewhat peculiar, president in Philadelphia. Yet, it was to trigger a fundamental change in the scientific system.

Back to Contents Page

Chapter 4

The science of science

> The science of science, or the self-consciousness of science, as I have put it elsewhere, is the real drastic advance of the second part of the twentieth century. (Bernal 1964)

4.1 Welcoming the *SCI*

At first, the *SCI* did not seem to have much impact on science. Its existence did not change scientists' information seeking behaviour. As has already been said in chapter 2, most of them seemed indifferent and the *SCI* failed to transform the system of scientific publication (chapter 3). The printed journal kept its position as the principal outlet of publication. Scientists as well as librarians and publishers stuck to the methods they were used to. As a consequence, it was more difficult to create a niche for the new bibliographic tool within the already existing matrix of social and cognitive relationships.

The *SCI* was, however, not only a bibliographic tool. Because of the series of decisions taken during its construction, it had grown into a large database which could be used in a variety of ways. Its broad coverage of all sorts of disciplines within science compensated in many respects for its shortcomings in journal selection. These *SCI* data appealed to two different groups: the sociologists of science on the one hand, and a motley collection of mainly natural scientists involved in "the science of science" on the other. The latter group especially saw the *SCI* as the long awaited instrument to study ánd steer science in an objective, quantitative, truely "scientific" way.

SCI's builders had been aware of its sociological and historical potential from the very beginning. Indeed, Garfield had stressed the historical uses of a citation index in his earliest publications (Garfield 1955) (chapter 2). This was not just a marketing strategy, it reflected his genuine interest in science. While constructing the index, he discovered many interesting facts in the data, reflecting interesting features of the scientific process. Together with his director of research Irving Sher, he published some of these findings on a fairly regular basis (Garfield & Sher 1966). When the final product came into sight in 1962, Garfield was quite prepared to draw scholarly attention to this instrument and data collection. He approached three key persons in the realm of the sociology of science: the sociologist of science Robert Merton, the physicist and historian Derek Price and the crystallographer John Desmond Bernal. In March 1962, Derek Price was advised of the upcoming *Science Citation Index*:

> There appears to be a great deal that the sociologist can do with citation indexes, but I would appreciate the opinion of an expert on this question. As I interpret comments that have been made along these lines, the citation index itself will not be so important as the research paths it will open for the sociologist who would otherwise be bogged down in the spade work needed to identify pertinent documents[1].

While Garfield emphasized the use of the *SCI* to locate documents (the *SCI* as a bibliographic tool), Price immediately saw the opportunity to get access to a huge collection of sociological data (the *SCI* as a bibliometric tool). He was thrilled by Garfield's letter: "I am strangely excited by the material you sent me on the Citation Index Project"[2]. Intensive correspondence followed, with Price constantly asking for more data (and seemingly getting most of it), as well as giving generous advice about the best way of setting up the *SCI*. Price wished to create a "calculus of science", modeled on econometrics and thermodynamics. Not that Price did not care for the more "soft" aspects of science — quite the contrary. Actually, Price went along with the term *science of science* only reluctantly. He would have preferred the term *humanities of science*, as he had called the field in his *Science since Babylon* (Price 1961). Throughout his entire scientific career he was on a quest for objective, solid, knowledge:

> I take the position that the workings of science in society show to a surprising degree the mechanistic and determinate qualities of science itself, and for this reason the quantitative social scientific investigation of science is rather more succesful and regular than other social scientific studies. It seems to me that one may have high hopes of an objective elucidation of the structure of the scientific research front, an automatic mapping of the fields in action, with their breakthroughs and their core researchers all evaluated and automatically signaled by citation analysis. (Price 1961, 194)

This may be called the defining position of the science of science movement. All proponents of the science of science saw the new bibliographic tool as an ideal instrument for the quantitative analysis of science.

In 1965, in an article in *Nature* Maurice Goldsmith explained the objectives of a new Science of Science Foundation, of which he was the first director: "By the science of science we mean the examination of the phenomenon of 'science' by the methods of science itself" (Goldsmith 1965, 10). This type of study would, among others, include: the sociology of science; the psychology of the scientist; operational research on science; the economics of science; and the analysis of the flow of scientific information, as well as the planning of science. Its truly reflexive nature was wedded to a strong quantitative orientation:

[1]Garfield to Price, March 6, 1962.
[2]Price to Garfield, March 15, 1962.

> The science of science begins to merit the name of science since, although some of the foregoing topics were investigated long ago, it is only now that quantitative treatments can be attempted.

The new foundation was a prestigious one, two members were even so famous that they scarcely needed an introduction[3] (Goldsmith 1965). At the occasion of the foundation's first annual lecture in that year, Derek de Solla Price spelt it out in more detail in his "The Scientific Foundations of Science Policy". Science policy should in his opinion be underpinned by the science of science:

> Let me emphasize that the universal need is for a *scientific* basis for knowledge about science and technology; the need is not primarily for a collection of policy statements, however wise, or for opinions of scientists, however well informed. (...) We need a special body of scientific knowledge which can be a basis for whatsoever policies, governments and citizens may request. Without such knowledge we might well flounder from one *ad hoc* decision to the next, and squander resources by adopting impossible ends or inefficient means. (Price 1965*b*)

One of science's key features, according to Price, is its cumulative structure. And precisely in analyzing this, the *SCI* would prove invaluable:

> To use another analogy, depicting the connexions between papers, the cumulating structure of science has a texture full of short-range connexions like knitting, whereas the texture of a humanistic field of scholarship is much more of a random network with any point being just as likely to be connected with any other. It happens that such considerations are most important in current research on the structure of scientific information systems, particularly in the new radical tool of citation indexing.

When Price wrote this, he was thorougly familiar with citation indexing. He had been aware of its possibilities and caveats since 1962. Yet, this may not have been the main reason for his appreciation of the *SCI*, those involved in the science of science who were not familiar with citation indexing reacted equally enthusiastically. For example, John Desmond Bernal, who had been informed by Garfield shortly before Price[4] but had not participated in the project, reviewed the *SCI* in spirited terms:

> The value of the *Science Citation Index* was immediately apparent to me because I had tried to do the same thing in reverse order in writing about various aspects of the history of science. (...) Its essential value is, as claimed to be, that it is a new dimension in indices which should enable the polydimensional graph on the progress of science to be mapped out for the first time. Such a graph is a necessary stage in drawing up or planning any strategy for scientific research as a whole. (Bernal 1965)

---

[3]"The Advisory Committee of the Science of Science Foundation at the moment consists of Lord Snow, Prof. Derek J. de Solla Price (Avalon Professor of the history of science and medicine, Yale University), Prof. J. D. Bernal, Prof. Asa Briggs (dean of social studies, University of Sussex), Dr. Alexander King (director for scientific affairs of O.E.C.D.), and Mr. Gerald Piel (publisher of *Scientific American*)."

[4]Garfield to Bernal, March 2, 1962. See for Bernal's reply: Bernal to Garfield, March 22, 1962.

4.2 Roots

The positive valuation of the *SCI* in the science of science movement was related to the movement's character. These researchers felt they were exploring unknown territory and they were eager for new, especially quantitative, instruments. Most of them thought they were the pioneers of this endeavour. They were seemingly unaware that they had been scooped by an active and prominent group of Polish philosophers cum sociologists. As early as 1923 an unsigned editorial in *Nauka Polska* had proclaimed the emergence of a new field, called *wiedza o nauce* or "knowledge about science"[5] [6]:

> Science, like other products of culture such as art and religion, is an object of research. The appearance of a separate "knowledge about science" stems to some extent from the needs of life. Stimuli of a practical nature, such as reflections on the present needs of science and the implied need for its planned support, and on the relationship of science to other areas of culture and social life, require further deliberation upon creative scientific activity and the conditions of its development. Wherever some activity has been undertaken, the need to establish its own theoretical foundation eventually makes itself felt; hence the need for *research on science*.

In their opinion, this research on science should be policy oriented, reflexive, empirical and pluriform. It looked into science as a cultural phenomenon. In 1925, the Polish sociologist Florian Znaniecki published a further programmatic statement on the movement, "The Subject Matter and Tasks of the Science of Knowledge"[7] According to Znaniecki, the modern theory of knowledge "possessing its own empirical properties and amenable to empirical investigation, acquires the traits of positive, comparative, generalizing, and empirical science. Thus it distinguishes itself from epistemology, logic and the strictly descriptive history of science"[8].

---

[5]The following account on the Polish science of science borrows heavily from Krauze et al. (1977).

[6]The full quotation is: "Contributions contained in the published volumes of *Nauka Polska* have brought into full relief a complex of problems pertaining to science as a social phenomenon; a separate division of the *knowledge about science* has been delineated. Science, like other products of culture such as art and religion, is an object of research. The appearance of a separate "knowledge about science" stems to some extent from the needs of life. Stimuli of a practical nature, such as reflections on the present needs of science and the implied need for its planned support, and on the relationship of science to other areas of culture and social life, require further deliberation upon creative scientific activity and the conditions of its development. Wherever some activity has been undertaken, the need to establish its own theoretical foundation eventually makes itself felt; hence the need for *research on science*. Thus the articles in *Nauka Polska*, though largely devoted to the needs of the separate sciences or to the organization of science, also include contributions devoted to problems of a theoretical nature, problems pertaining, for example, to the socalled *sociology of science*. The statements made above provide a guideline for the program and direction of the journal. Its main task is to draw society's attention to the topical, but insufficiently popular, problem of science; for this purpose it is desirable to isolate this problem from other related problems and to investigate it separately, from both the theoretical and practical points of view" (*Nauka Polska* 4:vii) (Krauze et al. 1977, 194–195).

[7]See on Znaniecki also (Merton 1941).

[8]Quoted in Krauze et al. (1977, 198).

By 1928, the terminology had developed from *wiedza o nauce* to *naukosnawstwo*, which may be translated as the science of science. Between 1928 and 1939 around forty meetings were organized where the science of science was discussed, often "well attended by distinguished representatives of the Polish intellectual elite" (Krauze et al. 1977, 196). The work of this group was mainly published in *Nauka Polska* until 1936 when the "science of science movement" published a new journal, *Organon*. The first issue was an English translation of the program of the movement, "The Science of Science", written by Maria Ossowska and Stanislaw Ossowski. They demarcated their use of the term "science of science" from the older German *Wissenschaftslehre* which denoted "logic understood in a very wide sense" (Ossowska & Ossowski 1936, 76). Their science of science was to be empirical through and through and would entail both "epistemological" and "anthropological" points of view, the first centering on the intellectual products of science and the second on scientific activities as part of culture. The Polish scholars divided the field into five different, partly overlapping, domains: the philosophy, psychology, and sociology of science, supplemented by problems of "a practical and organizing character" and the historical problems of science. An important assumption was the idea of the unity of science:

> There is only one scientific culture, absorbing all scientific achievements, wherever and by whomever they are attained. There are no competing scientific cultures, there are no competing sciences as there are competing religions or codes of law. All incongruity between various scientific theories is considered a provisional stage which has to be overcome in this or that direction. (Ossowska & Ossowski 1936, 81)

This global scientific culture was, according to the authors, the main cause of the rise of the science of science: "It was brought into existence not only by new interests, but above all by the *new reality*. As the discovery of electricity brought with it the creation of new divisions of physics, so the development of Science in modern society will bring about the rise of the science of Science" (Ossowska & Ossowski 1936, 82). The assault on Poland by the German nazis prevented the further development of this interesting intellectual tradition until after the Second World War. A somewhat different approach of science took over: the externalist marxism of the Russian physicist Boris Hessen. Ironically, though, his contribution triggered the development of essentially the same type of science of science as had been created by the Polish philosophers.

4.3 The science of science in Russia, the Ukraine, and the Soviet Union

4.3.1 Naukovedeniye

At the beginning of this century the Russians[9] were far ahead of the West in the history and sociology of science[10]. The geochemist, polymath and representative of the liberal Russian professoriate Vladimir Vernadskii started this tradition of teaching and research on the history of science in 1893. In 1902 he gave the first course on "the history of the modern scientific world view" at Moscow University. After 1917 he actively campaigned for his own institution: a campaign which led in 1921 to the world's first institute of the history of science and technology: the "Academy of Science Commission on the History of Knowledge". According to the historian Graham (1993), Vernadskii was convinced that he lived in the era of a third scientific revolution, in which the advent of relativity theory and quantum theory were the crucial events. In his opinion, history should be a form of self-study of science to increase the impact of these historic developments. Vernadskii's view of the development of science was sophisticated for its time, "even though contemporary historians of science would probably say that he overplayed the role of genius and the power of ideas, and underestimated the importance of social context and technology" (Graham 1993, 138).

In the twenties, a social study of science emerged which was called "naukovedeniye". Its profile is closely related to its Polish counterpart, comprising the sociology of science, science management, and science organization. It was both an attempt to improve the performance of scientific researchers and an attempt to better understand science as a social phenomenon. One of the first Russian scholars to use the term "naukovedeniye" was I. Borichevskii:

> On the one hand, it is a study of the inherent nature of science, a general theory of scientific cognition. On the other hand, it is a study of the social purpose of science, of its relations with other types of social creativity. It is something we could call the sociology of science. This area of knowledge does not yet exist; but it must exist: It is required by the very dignity of its object, i.e. of the revolutionary power of exact knowledge. (Quoted in Graham 1993, 151)

This new field was also promoted by the permanent secretary of the Academy of Sciences, S. F. Ol'denburg. In the twenties a solid block of research was carried out. Between 1921 and 1934, statistical and organizational surveys of Russian science were made that were "remarkable in their detail" (Graham 1993, 152). Moreover:

[9]The following is based on research done by Lyuba Gurjeva (Gurjeva 1992) in her capacity as research assistant to this project, as well as on the historical literature about science studies in Russia, the Ukraine and the SU.

[10]This prelude to Russian scientometrics is based on Graham (1993).

> Soviet authors in the twenties attempted to improve scientific research by suggesting changes in research techniques and the use of laboratory equipment; by proposing reforms in publication and indexing operations; by calling for information-retrieval systems, including primitive computers; and by developing quantitative criteria for evaluating the effectiveness of scientific research. (Graham 1993, 152)

In 1929, the Academy of Science was taken over by the Russian communist party, and Vernadskii's role came to an end. As head of the Commission on the History of Knowledge he was replaced by Nikolai Bukharin, one of the principal leaders of the Bolshevik Revolution. Two years later, this commission was converted to the Institute for the History of Science and Technology and Bukharin became its first director. Although he was not as involved in research as Vernadskii had been, Bukharin did not have a simplistic approach to science. As a political economist, he thought that science and technology were closely related to his own specialty as well as Marxist philosophy. In his view, all science is mediated by social, political and economic factors and cannot be separated from society. Moreover, he expected science and technology in socialist countries to differ substantially from their counterparts in capitalist countries. Bukharin enters this story because Stalin, in a political move to demonstrate his willingness to cooperate with "bourgeois intellectuals", ordered him to lead a top delegation of Russian scientists to the first international conference at hand (Werskey 1978). This happened to be the *History of Science Congress* in London in 1931. The Russians arrived completely unexpectedly and, rather spectacularly, by plane (See for this rather amusing history Goldsmith & Mackay 1964, Bukharin et al. 1931, Werskey 1978). The physicist Boris Hessen was among them, and he delivered one of the most famous speeches in the history of the history of science, a presentation that immediately created a sensation among historians of science (Goldsmith & Mackay 1964).

Hessen, much to the dismay of traditional science historians, distanced himself from their approach to Newton and his work:

> Thus the phenomenon of Newton is regarded as due to the kindness of divine providence, and the mighty impulse which his work gave to the development of science and technology is regarded as the result of his personal genius. In this lecture we present a radically different conception of Newton and his work. Our task will consist in applying the method of dialectical materialism and the conception of this historical process which Marx created, to an analysis of the genesis and development of Newton's work in connection with the period in which he lived and worked. (Bukharin et al. 1931, 151–152)

This had a tremendous impact[11]. Witness Joseph Needham's recollection:

[11] Although it must be said that this happened mostly by the printed version of the presentation. During the oral presentations at the conference, the Russians did not get the time they wanted, so it was agreed that they would distribute their presentations in English in the form of the booklet *Science at the Crossroads*. The Russian original texts would be published in 1933 (Vucinich 1982, Werskey 1978, 127).

> Perhaps the outstanding Russian contribution was that of Boris Hessen, who made a long and classical statement on the Marxist historiography of science, taking as his subject of analysis Isaac Newton. Here was a paradigm of the traditional history of science, so great a genius that he could not have been influenced by his environment at all, and certainly not by a subconscious appreciation of the needs of the society of the rising bourgeoisie of the 17th century. To suggest such a thing was, in terms of conventional thinking, almost a secrilegious act, in any case culpable of *lèse-majesté*. Yet this was the case Hessen worked out *in extenso*, tripping over proper names and making mistakes of detail on the way, but producing a veritable manifesto of the Marxist form of externalism in the history of science. (Bukharin et al. 1931, viii)

According to Vucinich (1982), two facts about Hessen's essay are "incontrovertible": "First, it was not only the first but also the last effort to cast a historical epoch in the development of science in the mold of classical Marxism. And, second, no serious historian of science in the West was ready to ignore Hessen's provocative mode of historical explanation" (Vucinich 1982, 128). Hessen's presentation influenced both John Desmond Bernal[12], "the father of the science of science"[13], as well as Robert Merton, "the founding father of the sociology of science" (Price 1965*b*)[14].

In short, Hessen brought the reputation of Newton, the great 17th century British scientist who had by the 1930s acquired God-like features, down to earth by stressing the social relationships shaping Newtonian physics.

Graham (1993) draws attention to a somewhat different interpretation of Hessen's address by focusing on its multi-layered nature. Whereas in the UK and the US, the "externalist" view of science attracted most attention, within the SU the speech was meant to defend relativity physics and quantum mechanics against politically inspired criticisms. Hessen tried to defend both physics and marxism:

> He believed that the development of twentieth-century physics could be analyzed in the same way that he explained Newtonian physics, and thought there was no more reason to accept attacks on materialism in the name of twentieth-century physics than there had been to accept such attacks in the name of Newton, whose religious views were merely a 'product of his time and class'. The unwritten final line was that when Einstein wrote on religion or philosophy he also merely expressed his social context and therefore these views should not be held against his physics. (Graham 1993, 150)

[12]For an extensive biography: Werskey (1978).

[13]See for example the back cover of Goldsmith & Mackay (1964): "It was J. D. Bernal who first foreshadowed, in *The Social Function of Science*, the need for 'a science of science' on which a strategy for research could be based.". This claim is probably based on the following statement in Price (1965*b*): "To my knowledge the first person to make any extensive inroads into the scientific analysis of science was J. D. Bernal in that monumental work of 1939."

[14]This episode shows that the rediscovery of the social context of the great icons of Western science did not emerge in the beginning of the 1960s, as Appleby, Hunt & Jacob (1994, 174) suggest, but already in the 1920s and 1930s.

According to Graham (1993), Hessen's paper is better understood as a result of this peculiar situation rather than as a model of Marxist analysis of science. Nevertheless, in the West the emphasis was on the latter aspect. As an unintended consequence, the specialty of science studies was effectively exported, though in its marxist form. As it transpired this was just in time. Stalinist purges subsequently effectively destroyed all social science, and murdered most of the members of the Russian delegation to the 1931 London congress[15].

4.3.2 Naukometria

Decades later in the sixties, the Soviet Union was forced to re-import the results of its earlier export. Bernal (1939) and Price (1961) exerted great influence, sparking renewed interest in naukovedeniye, made possible by Khrushchev's Thaw. Mikulinsky, Deputy Director of the Institute of the History of Science and Technology of the Academy of Science in Moscow, was an ardent supporter of the new "naukovedeniye" partly because he was convinced that Russia was lagging behind in science studies. The main proponents of quantitative science studies were G. Dobrov in the Ukraine and V. V. Nalimov in Moscow. They were two very different characters and gave rise to two different forms of scientometrics in the Soviet Union. Dobrov, who was the first to publish a book on the science of science in the SU (Dobrov 1966), was a party man whereas Nalimov had been prisoner in the Gulag[16]. Dobrov created an active and international scientific network and consistently promoted his style of science studies. Nalimov lost interest after a series of intensive seminars and did not actively promote himself. Rather he pursued a more abstract philosphy of language and science. For Dobrov the field of science studies was instrumental to and closely connected with science policy. In Nalimov's opinion, science was a self-organizing system and he distrusted politics and politicians. He thought that science studies should improve the science's self-understanding, not serve some instrumental goal. Nalimov stressed the need for an open scientific system, whereas Dobrov seems to have functioned pretty well in the closed shop system of scientific institutions in the SU. Both were steeped in information science, but in different ways. Their life stories correlate, though not in a causal way, with two different brands of scientometrics. As a consequence, from the sixties onwards, there was no longer just one type of quantitative science studies in the SU, but two differing styles.

4.3.3 The Moscow branch

Nalimov, born in 1910 into a Moscow professor's family, hit upon the quantitative study of science by accident:

[15] Hessen died in one of Stalin's concentration camps in 1938 at the age of 55. In 1978 he was fully rehabilitated (Vucinich 1982, 129).

[16] Nalimov was arrested in 1936 and spent the next 18 years in prison, amongst others in the Kolyma camp. As head of quality control in the metallurgic plant in Dzheskazgan (Kazakhstan), he used probabilistic methods. After World War II, he started to submit articles to the journal *Industrial Laboratory* (Nalimov 1992). Released in 1953, he started writing abstracts for the All-Union Institute of Scientific Information (VINITI).

> I got acquainted with scientometrics by chance. At the end of the 50's, I was working at the All-Union Institute of Scientific Information (VINITI) as an editor of the "Physics" journal. I worked at VINITI, because I could translate from three European languages. We were usually given articles for translating by specialists in languages who distributed the journals among the editors. And once I was given an article by Price and told that I probably was the only person who could translate it. It was in Italian. I liked that paper.
>
> It was devoted to the exponential growth of science. Together with Vladutz and Styashkin I wrote an article that established a link between cybernetics and information science. It was published in the "Uspekhi Fiziki" (Advances of Physics) (Vladutz, Nalimov & Stiazkhin 1959). The Director of the Institute Mikhailov was the first to react to the article: he invited my co-authors and reprimanded them severely. I was not a member of the communist party and the reproof did not affect me. The appeal to cybernetics was incriminating for two of my colleagues. At that time it was against the official ideology.[17]

It was only after being appointed at the Institute of Rare Metals in Moscow[18], that Nalimov was able to pursue his newly awakened interest in the quantitative study of science. When in 1966 the 1965 edition of the *SCI* arrived, Nalimov started an informal seminar on citation analysis. One of the nine members was I. M. Orient, an editor of the journal *Industrial Laboratory* in which Nalimov had been publishing from exile. She recollected the start of the seminar as follows:

> We began to work in 1966. Our crystallization center was V. V. Nalimov. Then he (at the University) received from Garfield the 1965 SCI. Then he invited people to learn to work with the SCI. At first we didn't treat it seriously but later we began to give a lot of time and effort to it. We "distributed" the SCI among ourselves. I took analytical chemistry. Vasilyev studied chemical physics, Granovsky - non-organic chemistry. I can't remember all now. Barinova - spectral analysis. Each had his own portion.[19]

The participants in the seminar, their "invisible college", shared Eugene Garfield's perception of the importance of citations: "Scientists almost always quote the work of their predecessors. (...) References have long served to practicioners an important means of information search. (...) Bibliographic references become a far more efficient means of information search if one follows them in reversed order" (Gilyarevsky, Multchenko, Terekhin & Cherny 1968, 32, 33). The *SCI* was used to search for information, to study the development of particular ideas, to evaluate these, and to study the general regularities of the distribution and aging of publications (Gilyarevsky et al. 1968, 38). The theoretical model they used however, was Nalimov's model of science as an information process.

[17] Interview with V. V. Nalimov by Lyuba Gurjeva, December 11, 1991, Department of Biology, Moscow State University, Moscow.

[18] Nalimov joined the newly created Indepartment Laboratory of Statistical Methods at the Moscow State University on invitation of Kolmogorov, whose deputy he became.

[19] Interview with I. M. Orient by Lyuba Gurjeva, 21 January 1992, Moscow.

Nalimov was the acknowledged central figure: "He was our leader. (...) I was driven by the interest to the work of professor Nalimov"[20]; The work was done by hand, of course. Since many Russian-language journals were not included in the *SCI*, these early scientometricians had to collect a lot of additional information by themselves. According to the interviews, the nature of the group was that of a movement or a hobby club. "They spent their free time ruffling through journals, searching for references and counting publications and citations. Many of them began to wear glasses - the printing quality of the *SCI* was not very high. they called it 'experimental work', a close analogy with the laboratory research they were conducting in physics or chemistry." (Gurjeva 1992).

The citation analysis seminar led to a large collective publication in 1967 (Barinova et al. 1967), the summary of a year's work:

> After each of us completed his work in 1966 and contributed our works into that article of twelve authors. Nalimov wrote it and we received nearly a 1,000 rubles for it and went to a restaurant. We continued to work experimentally and come out with our own ideas. Nalimov didn't give us ideas, he only generalized our results. Neither has he ever given anybody a theme. All of us were specialists of a very high level. Each knew his subject very well.[21]

The collective publication concluded that "the series of studies had established a modest influence of Soviet science on world information flows" (Barinova et al. 1967). In other words, either Soviet science did not have a detectable impact because of delays or the international flow of information was not linked up with the SU flow of scientific information[22]. After this publication, the seminar moved to the Institute of the History of Science and Technology and acquired the status of a continuing seminar on quantitative methods of studying the development of science. These attracted wider attention, were published, and became part of the reports and memoranda of the IHST to the State Committee on Science and Technology and to the Presidium of the Academy of Sciences[23]. One year later, Nalimov published a small book in which he coined the term "naukometria" (scientometrics)[24], which had not been used previously by anyone else (Nalimov & Multchenko 1969). At the same time, Nalimov introduced the new specialty to the All-Union Conference *Dokumentalistika-69*, in a section titled "New Ideas in Scientific Information". Following Garfield and Price, Nalimov argued that a new task had arisen for documentation science:

> At the present time one can confidently speak of the emergence of a new field of knowledge: science studies. (...) Science is a process actually existing

[20] Interview with Youri Vasilyevitch Granovsky by Lyuba Gurjeva, 5 November 1991, Chemical Department of Moscow State University, Moscow.

[21] Interview with I. M. Orient by Lyuba Gurjeva, 21 January 1992, Moscow.

[22] As has been discussed in (chapter 2) this had also been a major concern in the US a decade earlier.

[23] Scientific Archive of the IHST, F. 2, Op. 2, Ed. khr. 20, 1.5.

[24] Originally, this was translated by the American Rand Corporation as *sciencemetrics* (Nalimov 1966).

> and developing over time and characterized by definite quantitative parameters, and it is only natural to attempt to study it in the same way that natural scientists study the chemical, physical and biological processes developing over time. (...) It is natural to term the quantitative methods of studying the development process of science scientometrics. (Nalimov 1966, 1).

Nalimov approached science as an information process in which citations could be used as indicators:

> Science is a system. What makes it a system? It is not a chaotic flow of publications! It has structure. It is structured by citations. Citations show, how separate publications are related to one another, how science is developing as a result of interaction of publications in a particular domain and even in several domains. Thus what citation characterizes is not the achievement of a separate scientist but a contribution to the information process.[25]

The book was not published without difficulty:

> My small book was published with great difficulty. It was only thanks to the effort of one of the editors of the Nauka publishing house that the book was published. She concealed it from the chief editor. When it was published, they got a coded cable from the Central Committee of the communist party of the German Democratic Republic. The book was called intolerable. It turned out that they had planned to translate it. One of the members of the Central Committee was to edit it. And he discovered that in the tables it was written: France, USA, ...German language... There was no one Germany then. Because when we could not attribute a paper to one of the two German States, we wrote: "German language". They urged that the papers be divided between the two countries. The Soviet editor received a reproof. This is how it was developing. Now it may seem funny. But there has always been a resistance.[26]

Nalimov thought that this resistance arose from two sources: an ideological resistance to the idea that science itself could be analyzed quantitatively, and the fact that "citation analysis yielded unpleasant information, especially for some persons"[27].

The main audience for Nalimov's group were not science policy officials but scientists. After 1969, Nalimov's interest turned to other research topics and the group was disbanded. Most of the scientometricians pursued their research, but individually, dispersed over various institutions and without many resources (Gurjeva 1992).

[25] Interview with V. V. Nalimov by Lyuba Gurjeva, December 11, 1991, Department of Biology, Moscow State University, Moscow.

[26] Interview with V. V. Nalimov by Lyuba Gurjeva, December 11, 1991, Department of Biology, Moscow State University, Moscow.

[27] Interview with V. V. Nalimov by Lyuba Gurjeva, December 11, 1991, Department of Biology, Moscow State University, Moscow.

4.3.4 The Kiev branch

Compared with V. Nalimov, G. Dobrov was a more succesful builder of social institutions. Born in 1929 in the Donetsk Bassin, he was interested in the history of technology from the very beginning[28]. He became an active and rather succesful Komsomol member, rose to the rank of regional secretary of the Komosomol, but then decided to go back to academic research — not the usual career pattern. In 1961, he became head of the department of History of Technology at the Institute of History of the Ukranian Academy of Science. Three years later, he created the "Lab of Computer Methods of Processing Information on the History of Science". Quantitative methods in social science were popular at that time being, as they were related to the fame of cybernetics[29]. Dobrov, like some other Soviet authors, disagreed with Derek Price on the levelling off of the exponential growth curves, arguing that computers would enable a continuing rise of productivity in science[30].

Like Derek Price, Dobrov wanted to reveal regularities in the development of science. He saw this as the basis for evaluating and regulating scientific and technological activity, as well as for preparing and informing science policy. He thought that mathematical methods would provide the means of forecasting the development of science and technology (Dobrov 1964). In 1966, Dobrov published his influential book "The science of science" (Dobrov 1966).

The Kiev branch of scientometrics was, contrary to its Moscow counterpart, strongly policy-oriented. After 1968, forecasting scientific and technological development became an important political topic in the SU and Dobrov had time serial data. Dobrov moved into the field of forecasting science and technology. He led the "Division of the Complex Problems of Naukovedeniye" of the Council for Studying the Productive Forces, which became part of the Institute of Cybernetics of the Academy of Science in the Ukraine in 1971. This intimate relationship with a government body would remain a crucial feature of Dobrov's scientometric institution in Kiev. It was the centre of a specialist Ukrainian journal *Naukovedeniye i Informatika* (Science Studies and Information Science) and trained graduate students. The centre organized a host of national symposia from 1966 on. In the seventies they went international, attracting among others Eugene Garfield and Derek Price to Kiev. These symposia were effective network builders and raised the international standing of their founder Dobrov. In fact, he became one of the best-known scientometricians in the world, working at a host of international institutions (amongst which IIASA in Laxenburg, Austria) and being appointed as an expert to UNESCO, as well as an advisor to the Iraqi government on science policy matters.

[28] Interview with A. Korennoy by Lyuba Gurjeva, 16 October 1991, Centre for Studies of Science Potential and History of the Ukraina AS, Kiev.

[29] The popularity of cybernetics fluctuated rather strongly in the SU.

[30] The reception of Derek Price's work in the SU was mixed. Whereas Mikulinsky and Nalimov agreed with him, Dobrov and other authors were rather critical. The main point of criticism was that Price tried to find the mechanisms of the development of science within itself, whereas these authors claimed priority for "the demands of public production and social practice on the whole" (Gurjeva 1992).

Yet, it is highly questionable whether this flurry of scientometric research had any impact on SU science policy at all. For example, a crucial Ukrainian scientometric project was the building of ISTOK, the information system of thematic orientation and classification. Its central technique was shaped on the co-citation technique (section 5.4.3, page 116) and transformed to deal with technical reports: "the co-request method". Dobrov's PhD student A. A. Korennoy used the data of the All-Union Centre of Scientific and Technical Information[31] to trace the patterns in the request for research reports:

> If Small, Garfield and Marshakova used citation data, we used non-published reports and as indicators of connection - not citations but requests of these reports. The results were maps of research done in the country in various fields. The advantage of such approach is the most recent information. Report can be ordered in month after the research is finished, while one has to wait for 1,5-2 years for a book. The disadvantage is that this system is closed. A copy of the report can be ordered only by the institute. An individual scientist can not do it.[32]

This partially closed character also determined the scientometric research products of the institute. Most of them were reports for governing bodies in the Ukraine, prepared at the request of the State Committee on Science and Technology, and of the Presidium of the Academy of Sciences in Kiev. The effect they subsequently had on science policy is unclear to say the least:

> It is a disease not only of scientometrics and science studies but of our all our science. When we proceed from theory to practice, the effectiveness of our research falls, sometimes down to nil. It means that all our work, although we finished every project with the report on its implementation and the yielded economic effects, is on paper. The reasons for that are, first of all, the absence of interest on the part of the government and its bodies, Presidium of the State Committee on Science and Technology (SCST) - our clients. I mean that they have no practical need for the results of scientometric studies. We thought that as we put the map of some branch of science on top of the desk of the head of the State Committee, it becomes clear to him who is the leader in this sphere, how the research front is to be altered, who is to be financed - and theoretically it was correct. But actually no one in our whole system of management ever needed that kind of documents. The decisions taken were mostly voluntaristic and guided by completely different considerations. The funds were allocated not according to the front of research but according to personal acquaintanceship. Under such conditions our works had a character of demonstration. Here we have some results, look how interesting it is, how nice! You can use it ... We have also got some recommendations. After that we got a certificate proving the utility of our results, their potential economic effect and everyone was happy. We were financed and had an opportunity to write our monographs and dissertations. And there

[31] Unlike VINITI that dealt with international publications, this institution accumulated research reports of Soviet institutions.

[32] Interview with A. A. Korennoy by Lyuba Gurjeva, 16 October 1991, Centre for Studies of Science Potential and History of the Ukraine, the Academy of Science, Kiev.

> was the impression of scientific activity. There was a seeming interest on the part of the client but actually the results of most of our work were piled on the shelves.[33]

4.3.5 Two different styles

In summary, the differences between the two scientometric schools in the former SU are found in the concept of science, the audience of the scientometricians, the objects they used, the way they organized their research and their personal scientific style. Although both Nalimov and Dobrov were steeped in information science, their conceptualization of the character of science differs substantially. Nalimov emphasized the self-organizing character of science as an information process, in which the possibility of steering by policy were only marginal. On the contrary, Dobrov defined science as a cybernetic organism or instrument, greatly influenced by science policy. This correlates with the different audiences the schools addressed. For Nalimov's "invisible college", practicing scientists were the main public. For Dobrov's institute, science policy bodies were the principal clients. Moreover, the objects of study differed somewhat. Nalimov focused on published articles in peer-reviewed scientific journals, whereas Dobrov made use of the technical "grey literature". Lastly, their personal styles were quite different. Nalimov was a philosopher who flourished in a small network of closely collaborating colleagues, their own "invisible college", but does not seem to have been very interested in creating durable institutions. Dobrov, on the other hand, was a supreme builder of institutions, policy related visible networks and international fame.

4.4 Western science of science

John Desmond Bernal, who seems to have been unaware of his Polish and Russian predecessors[34], decided after the 1931 *History of Science Congress* to write along similar lines as Hessen's analysis, which resulted in his 1939 masterpiece (Bernal 1939). Anxiety about the abuse of science for "base ends" (Bernal 1939, xv) and the "appalling inefficiency both as to its internal organization and as to the means of application to problems of production or of welfare" (Bernal 1939, xiii) provided the main perspective of this book. This was summed up in the phrase "the frustration of science"[35]: "For those who have once seen it, the frustration of science is a very bitter thing" (Bernal 1939, xv). He saw science as the solution to many problems of humanity, if not all, if only the social relationships could prevent its abuses. Therefore, he was in favour of submitting science to communism:

[33] Interview with A. A. Korennoy by Lyuba Gurjeva, 16 October 1991, Centre for Studies of Science Potential and History of the Ukraine, the Academy of Science, Kiev.

[34] That is, if we may take the absence of any reference to the Polish works in Bernal's *Social Function of Science*, whether implicit or explicit, as indicative.

[35] This phrase was not invented by Bernal, it has older roots. For its use by the Dutch see Molenaar (1994).

> The relevance of Marxism to science is that it removes it from its imagined position of complete detachment and shows it as a part, but a critically important part, of economic and social development. (...) Already we have in the practice of science the prototype for all human common action. The task which the scientists have undertaken - the understanding and control of nature and of man himself - is merely the conscious expression of the task of human society. The methods by which this task is attempted, however imperfectly they are realized, are the methods by which humanity is most likely to secure its own future. In its endeavour, science is communism. (Bernal 1939, 415)

Interestingly, the term "science of science" is not used by Bernal (1939). This is not to say that it is nonsense to see Bernal as one of the founding fathers of the science of science. On the contrary, his book gave a huge boost to the idea of scientifically analyzing science in a period when the earlier, and probably somewhat more subtle, Polish voice was stifled. Moreover, Bernal also presented an attempt to quantify the state of science[36]. He stressed the need of a scientific rationale for science policy, the reflexive[37] systematic application of scientific methods to science itself, and the idea of an overall unity of science.

As has been noted in chapter 2, all main actors in this study of scientometrics were influenced by his ideas in several respects: Eugene Garfield, Joshua Lederberg, Robert Merton, Derek Price, and V. V. Nalimov. Derek Price was central to the science of science movement. However, he did not share Bernal's ideological commitments. As Price told his audience at the first annual lecture of the Science of Science Foundation:

> This was only a couple of years after J. D. Bernal had published his book on *The Social Function of Science*, which introduced many of us to the several types of issues involved, and proposed a number of radical solutions. In the great "Planning of Science" debate which was then raging we heard such issues discussed, often by able scientists, but almost always as if they were political or doctrinal questions to be argued from inner convictions and on the basis of obvious truths about science - over which they, however, seemed to disagree heatedly among themselves. (Price 1965*b*, 233)

To Price the scientific method, and not socialism, was central: "It is perhaps especially perverse of the historian of science to remain purely a historian and fail to bring the powers of science to bear upon the problems of its own structure. There should be much scope for a scientific attack on science's own internal problems" (Price 1961, 162). The historian of science was privileged because of science's special characteristics[38]:

[36] This is done in appendix VIII of Bernal (1939).

[37] In Bernal's reasoning this reflexive character relied heavily upon the Marxian and Hegelian notion of "self-consciousness".

[38] Price's program was to find simple structures underlying more complex surface phenomena of science. This has become a common feature of scientometrics. Cf. Leydesdorff (1986, 8): "De overweldigende complexiteit van sociale verschijnselen maakt ons soms huiverig om over deze problemen in analytische termen na te denken. Het lijkt alsof in de historische ontwikke-

> Now the history of science differs remarkably from all other branches of history, being singled out by virtue of its much more orderly array of material and also by the objective criteria which exist for the facts of science but not necessarily for the facts of other history. Thus, we can be reasonably sure what sort of things must have been observed by Boyle or Galileo or Harvey, in a way that we can never be sure of the details of Shakespeare's life and work. Also, we can speak certainly about the interrelations of physics, chemistry, and biology, but not so positively about the interdependence of the histories of Britain, France, and America. (Price 1961, 162)

Reflexive quantification fascinated Derek Price from the very beginning. In 1965, he recalled the beginning of his craze for measuring science as follows:

> In 1942 I heard Sir Lawrence Bragg make a statement, in the Royal Institution, which seemed curiously intriguing. In the course of a discussion on the training of physicists he remarked that Britain was producing about one good physicist per million of the population per annum. As a physicist newly graduated that year, I suppose there was room for me to feel personally involved with the rarity of the commodity to which I aspired, but more what caught my fancy was the sudden recognition that we could think about physics with the same powerful methods that we use to think in physics itself. If one could make numerical guesses about orders of magnitude, it could lead to accurate quantitative theory, and if that were possible one might explan qualitative phenomena with all the power and objectivity of science. (Price 1965*a*, 233)

Price presented his first paper on the history of science at the sixth *Congrès International d'Histoire des Sciences* in 1950 in Amsterdam. It dealt with quantitative measures of the development of science (Price 1951) but as he noted himself "it fell flat". He did not get any response whatsoever. The main cause of this was probably that his audience was unfamiliar with, or even resentful of, the use of quantitative methods in history writing[39]. In Price's first historical paper, the policy orientation commonly seen in the science of science can be traced as well. The study of science should, according to Price: "enable one to diagnose the nature of changes currently taking place, and to plan accordingly in the disposition of research facilities at university and other laboratories" (Price 1951, 92). This attitude — which together with Price's quantitative inclination may have contributed to the failure of his first paper — made his views particularly attractive in the later policy debates of the early sixties in the United States. His Pengram Lectures in 1962, published as *Little Science, Big Science*, attracted the attention of most science policy makers. As one of them later recalled "we all had this book on our night-stands".

ling 'alles met alles samenhangt' en dat met name de ene ontwikkeling de andere oproept en dat beide elkaar determineren. Toch liggen er soms nogal eenvoudige structuren ten grondslag aan wat heel complex lijkt, zoals we in het begin van dit hoofdstuk voor bijvoorbeeld publicatiepatronen hebben aangegeven. Het is het programma van het kwantitatieve wetenschapsonderzoek om die structuren op het spoor te komen, en die bij voorkeur in eenvoudige, door de computer hanteerbare modellen te vatten."

[39] Appleby et al. (1994, 86-86) link the use of quantitative methods in history writing to the *Annales* school and the American tradition of modernization theory.

4.5 "Please reply with more data"

The work of Derek de Solla Price embodies the convergence of the science of science and the vast resources opened up by the *Science Citation Index*. As has already been mentioned, Price instantly realized the potential of this database. He visited Garfield in January 1963[40] to discuss possible forms of collaboration. Price was quite eager for data: "Please reply with more data"[41] could have been the motto of all his letters to Garfield. Garfield did not always feel he could satisfy Price's appetite for data, especially since the contracts with NSF and NIH did not give much leeway to divert time and money into side-issues[42]. Price also encouraged Garfield to pursue research for himself. For example, in 1964 he wrote:

> I am most excited about your paper on "Statistical Analyses of International Chemical Research by Individual Chemists, Languages and Countries". It is full of the most important data that may well be crucial in our analyses. I congratulate you warmly on a very succesful job and urge you most strongly to be bold and publish it. I think it is most important that you take your courage in both hands and behave like a Hollywood scientist who must publish truth whatsoever the danger. I have no doubt that several chemists will be deeply hurt, but I cannot think it will be entirely bad for business.[43]

This cooperation seems to have been mutually advantageous. Since Garfield was not often acknowledged as a social scientist[44], his cooperation with Price increased ISI's academic credentials. To Price the main bonus was access to an seemingly unending stream of data with which he used to toy and produce his now famous research projects, papers and seminars.

This playful yet serious work with data has formed part of scientometric's identity. Its central motive was the discovery of simple and general laws governing science, comparable for example to the physicist's search for the elementary building blocks of matter. Much of the work consisted of descriptive statistics, like the various exponential growth curves or "Garfield's constant"[45]. The research focussed on science as a whole, and on its being demarcated from both the social sciences and the humanities. Price and his collaborators were interested in the question of whether these intuitive distinctions could be measured more accurately. For example, in 1966 Price compared the time distribution of articles in the 1965 *SCI* files with those in the 1961 files. His most important conclusion from

[40]Garfield to Price, January 31, 1963.

[41]Price to Garfield, February 7, 1963.

[42]Garfield to Price, February 14, 1963.

[43]Price to Garfield, 15 December 1964.

[44]For example, an application to the Sociology and Social Psychology program of the National Science Foundation was turned down as follows: "If a qualified sociologist of science were to propose a specific research project which involved the use of your citation indexes, we would be happy to consider such a research proposal and to include a revision of your dictionaries to suit the research objectives of the principal investigator. Without such a specific research problem to justify and guide the form of the revision, we would regard its value as questionable." (Hall to Garfield, May 2, 1963).

[45]This was the relationship between the number of documents and the number of references in the *SCI* files. Part of the discussion focussed on the question whether this actually was a constant.

this was that "the comparison of the two years confirms my conjecture about the existence of a research front"[46]. The goal of this work was to discern the "really existing" structure in science, to be gauged by clustering the outcomes of citation analysis:

> I had a session last week at the National Institutes of Health and managed to get across to them the idea that we can probably use citation network techniques to generate fields of scientific activity. Such fields produced by cycling[47] in the index are related to people and papers and not to arbitrary index classification[48]. I suspect we can show the "real" existence of these fields and all fields, and show whether a paper is central or peripheral within them.[49]

Almost twenty years later, Price was still eager to reveal the structure of science. Witness his reaction to ISI's Atlas of Science project (Starchild et al. 1981, Garfield et al. 1984):

> the Atlas project seems to my mind to be blossoming into the fairly certain probability that one can decompose the entire SCI corpus into something like 5000 separate clusters each of which constitutes an isolatable sub-field and invisible college. Very few papers fall in these interstices between clusters and the overlap between contiguous clustes is rather small.[50]

As will be clear from the above, this type of scientometric work was partly oriented to science policy. Price and colleagues were keen on applying their findings to science policy or elucidating controversial issues. The science of science was not a goal in itself, but was meant to be instrumental for science as a whole.

4.6 The citation sociologically used

The sociology of science differs from the science of science movement in the way it positions itself and in the way it uses the *SCI*. In 1939, Bernal had not been the first to present a quantitative analysis of science. Four years before, Robert Merton had finished his thesis *Science, Technology and Society in Seventeenth-Century England*, which was published in 1938 in *Osiris* (Merton 1938). Merton's thesis has a strong quantitative character. It has become known mainly because of its analysis of the interrelation between Puritanism and the institutionalization of science[51]. However, there is much more to it, as Merton underlines himself:

[46] Price to Sher, July 14, 1966.

[47] Cycling is a specific technique of accessing the data in the *SCI*: one starts by collecting citations, then looks up the references in those papers, collects citations to these references, and so on. In this way, a large collection of related papers is harvested.

[48] Price is probably referring to subject indexing.

[49] Price to Garfield, November 6, 1963.

[50] Price to Garfield, June 2, 1981.

[51] See for example Vucinich (1982, 129): "he concentrated primarily on the Protestant ethic as an independent variable in the whirlpool of seventeenth century history".

> somewhat more space in the monograph has been devoted to the hypotheses about economic and military influences on the range of scientific inquiry than to the hypotheses that link up Puritanism with recruitment and commitment to work in science. And yet, as I have told, the trio of chapters on the second subject has received all manner of attention in scholarly print while the quartet of chapters on the first subject has received remarkably little. (Merton 1978, xiii)

Merton would like to have it the other way around, not in the last place because the section on economical and military influences rejects:

> the mock choice between a vulgar Marxism and an equally vulgar purism. The need for this evenhanded rejection of both simplisms is widely recognized today. But when the monograph was written - during the Great Depression, it will be remembered - vulgar Marxism was just about the only variety of Marxism that was being expounded on the periphery of American academic circles. (Merton 1978, xiii)

Notwithstanding this rejection of vulgar marxism, Boris Hessen's approach clearly impressed the young Merton. Witness the following footnote in that important chapter on economic influences on science:

> In the discussion of the technical and scientific problems raised by certain economic developments, I follow closely the technical analysis of Professor B. Hessen in his provocative essay (...) Professor Hessen's procedure, if carefully checked, provides a very useful basis for determining empirically the relations between economic and scientific development. These relations are probably different in an other than capitalistic economy since the rationalization which permeates capitalism stimulates the development of scientific technology." (Merton 1978, 142, n. 24)

Merton's thesis is a classic in the sociology of science. It uses quantitative approaches to elucidate scientific developments. In this respect, various older intellectual traditions made their mark in Merton's book: the quantitative inclination of Merton's teachers George Sarton and Pitirim Sorokin, as well as the nineteenth century traditions of Francis Galton (Galton 1874) and Alphonse de Candolle (de Candolle 1885). Moreover, Merton was aware of the older bibliometric tradition of citation analysis among librarians[52]. This sensitivity to quantitative instruments goes a long way to explain the warm reception Merton gave decades later to the *SCI*.

Garfield informed Merton about the citation indexing project a few days before Price[53]. He responded, like Price, quite positively:

> After having read the offprints you were good enough to send me, I am persuaded that your materials should be a rich source for the sociologist of science. As it happens, I am now in the midst of working on a problem in this field which needs precisely the kind of evidence you are putting together in your Citation Index[54].

[52]For example, Merton refers to Cole & Eales (1917) (Merton 1978, 4, n. 7).
[53]E. Garfield to R. K. Merton, March 1, 1962.
[54]R. K. Merton to E. Garfield, April 19, 1962.

However, contrary to the interaction between Garfield and Price, Merton's reply did initially not lead to collaboration on research projects. Merton and Garfield did not even meet before a meeting of the *American Association for the Advancement of Science* in 1969[55]. The *SCI* made its way to the Columbian University sociology of science mainly via Merton's research assistants. One or two years after Garfield's letter to Merton, Harriet Zuckerman, who was doing her PhD with Merton, was informed about the *SCI*:

> The first contact I had was in 1963 or 1964 when Gene had a sort of meeting in New York for people that might possibly be interested in the science citation index. It was in a hotel somewhere and he asked Bob to come. And Bob didn't want to take the time to go but it looked interesting and so he asked me. I was working on my dissertation. It was a presentation of what the *SCI* produced. So I knew about citation indexes really very early, though I cannot say that I perceived the usefulness of citation indexes for doing research in the sociology and historical sociology of science until somewhat later. But it is evident at least that we understood from very early on that citation counting could be used as a rough indicator of the impact of scientists' work. And I think that we also understood that it would be very useful in tracing up the affiliation of ideas. But we obviously didn't have the sense of what was going to come.[56]

In 1966, Merton and Zuckerman started an advanced seminar (for graduate students) in the sociology and history of science which attempted to use the *SCI* as a research instrument. "What we tried to do in the seminar was to develop a method, which I thought we needed badly, to identify the substance of problems emerging in the history and sociology of science. We were not focused on the metrics but on the substance of the problems" [57]. The work done in the seminar was unconnected with the research done by Garfield and his colleagues.

> The two Cole brothers were members of the seminar and it was in that seminar that we talked an awful lot about the wonderful things that could be done with the citation index and about the pitfalls. And I think it was in that seminar that we really explored the diversity of things that could be done, and Jon (Cole) and Steve (Cole) worked very hard doing their particular type of analysis. It was not through Garfield.[58]

In fact, the sociologists thought the ISI papers lacked methodological sophistication: "I remember a paper by Sher and Garfield and that struck me as being, you know, useful but highly, highly limited. Not just in the data that was presented but limited in the methodological development. So I would say that what was going on at Columbia and what was going on in Philadelphia were basically unconnected"[59]. The Columbian research program had been inspired by a 1957

[55]E. Garfield to R. K. Merton, March 25, 1969.

[56]Interview with Harriet Zuckerman, 2 December 1993, New York.

[57]Interview with Robert Merton, 2 December 1993, New York.

[58]Interview with Harriet Zuckerman, 2 December 1993, New York.

[59]Interview with Harriet Zuckerman, 2 December 1993, New York.

paper by Merton on the reward system in science and its initial focus was on the way scientists were credited (among others by citations) and the relationship between reward and actual contribution. This was a stronger drive than the policy related orientation and the interest in the natural sciences themselves in the science of science. For Steve Cole, for example, science was simply a resourceful research site:

> The only way I ever got to the sociology of science was pure luck. It simply had to do with what Merton was interested in at the time. If somebody had told me while I was an undergraduate at Columbia that I would become a sociologist of science, I would have said they were crazy. Cause I was never particularly interested in science per se. (...) It turned out, however, that science was a particularly good field for me because it was so easy to get data. That is I think the biggest problem in sociology. In science everything is published! It is all public record.[60]

At the seminar such problems were explored as the way of recognition; the question of whether a person who is in a place outside of the mainstream is recognized as much and as quickly as those who are central; the impact of of getting famous on subsequent citation. It was all a way of filling out the theoretical propositions that Merton had made earlier on. Sometimes this created tensions, especially if the empirical results seemed to contradict the theoretical expectations': "There was quite a bit of discussion constantly during the whole colony of sociology of science between Merton and particularly myself debating what you could conclude on the basis of these empirical studies. Part of the problem was that findings didn't show support for some of his ideas"[61].

The seminar had its own dynamics. Zuckerman, the Cole brothers and their colleagues began to see more dimensions of the citation index and more ways it could be put to use (Cole & Cole 1971, Cole 1970, Cole & Cole 1967, Zuckerman 1971). "We began to see that it had potential for understanding the development of knowledge and the extent to which sciences impinge on one another and were interconnected. We saw that the tool could be very useful in a lot of different ways including in studying what we at that time began to call the cognitive structure of science. I think that we knew that we were doing something different from what we had done earlier, but I don't know if we had a handy way to talk about it"[62]. It was a critical enthusiasm, though, the awareness of pitfalls, problems and shortcomings of the *SCI* as a sociological tool was always present:

> All through this period, as a kind of underlying second programme of research was the effort to understand what the shortcoming and problems were of citation analysis. That is what we were always worried about. We were always aware of that because we talked a lot with scientists. (Some sociologists don't and some do, we were in the crowd of those who did.) And some of them were always telling us what was wrong with citation counting.

[60]Interview with Stephen Cole, November 1994, New Orleans.
[61]Interview with Stephen Cole, November 1994, New Orleans.
[62]Interview with Harriet Zuckerman, 2 December 1993, the American, New York.

> They were determined to keep us honest, so to speak, in using citation analysis. So we were always trying to see ways in which we could deal with the kind of criticism they made. For example: it is very common for scientists to say that citation counts don't really measure the importance of a scientist's contribution because there are a lot of negative citations. Well in that same seminar some years later, Suzan Cozzens began to look at the context and content of citations.[63]

This critical attitude laid the foundation for a more intensive communication between the citation indexers in Philadelphia and the sociologists at Columbia. Merton and Garfield finally met in 1969, seven years after their first communication, at the annual meeting of the American Association for the Advancement of Science in Dallas. Immediately after this meeting, Garfield invited Merton to join the board of the *SCI* and the new social science edition of *Current Contents*[64]. Merton preferred to take up the first opportunity, since "I agree with you that it provides a wealth of pertinent information for sociologists of science"[65]. Harriet Zuckerman, then Assistant Professor of Sociology at Columbia University, happened to be studying Nobel prize winners, a topic about which Garfield had given a presentation in Dallas. Merton had passed the paper to Zuckerman, who in turn sent Garfield her paper on the topic[66]. Two years later[67], Zuckerman requested citation data "for work I am now completing on Nobel laureates"[68]. In his reply, Garfield spelled out his new policy[69] with regard to giving scholars the *SCI* data:

> we have established a policy that we believe is reasonable to an ethical handling of the authorship problem. (1) If small amounts of data are to be quoted or used, as above, we expect proper reference to our work. (2) If the amount of data is moderate, then we expect appropriate recognition of its origin. (3) Finally, if the data used are extensive, as implied by your letter, then we <u>may</u> expect to participate in the authorship. In all cases, we must insist upon a pre-publication opportunity of judging whether or not the data are interpreted and used correctly.[70]

One of the particular problems in correctly interpreting the data related to *SCI*'s only publishing the name of the first author, and to the persistent problem of people having the same name. Zuckerman withdrew her request for the data[71].

[63] Interview with Harriet Zuckerman, 2 December 1993, New York.

[64] Garfield to Merton, 25 March 1969. Derek Price was also member of *SCI*'s advisory board.

[65] Merton to Garfield, 26 March 1969.

[66] Zuckerman to Garfield, 22 April 1969.

[67] In the mean time, Zuckerman had requested a reprint of Garfield & Sher (1966).

[68] Zuckerman to Garfield, June 22, 1971.

[69] This was instigated after ISI was asked for data on a bigger scale.

[70] Zuckerman to Garfield, July 16, 1971.

[71] "I can well understand your concern about making data available on which you or your staff members are now at work. For this reason, I withdraw my request for a copy of the 100 most cited authors and articles. The only reason I wished to have it was to see how frequently Nobelists and their counterparts turn up. I was not interested in publishing the list itself." (Zuckerman to Garfield, July 27, 1971).

She provided Garfield with a paper on the way scientists decide upon the order of the author names (Zuckerman 1971), indicating that this depends on the scientific field, the age and the eminence of the authors involved. Zuckerman drew a critical conclusion:

> Since you have observed that citations are an indication of scientists' impact on their fields, these patterns of authorship are of great significance. (...) The present system of using first authors only in the SCI should be changed. First, a large majority of multi-authored papers have two authors. If two authors' names were listed instead of one, full authorship would be provided for much of the literature. Second, for larger author sets, it might be helpful to list the names of first and last authors, since last position is important in the life sciences especially.[72]

Garfield subsequently explained that Zuckerman's proposal was

> completely impractical for the present. In fact it would be a complete reversal of the main reason why the <u>Citation Index</u> was created in the first place. It was not created as a tool for measuring the performance of scientists! It was, and remains, primarily a tool for information retrieval. As such, the first author, year, journal, volume and page is more than sufficient as an identifier[73]

In November of that year, Garfield paid Zuckerman a visit and showed her colleagues how to use the *SCI*[74]. A year later, Zuckerman requested data for a comparative study of the sciences for the first time[75]. The Columbia University group was far less data-driven and attached more merit to theoretically underpinned research designs than to the more descriptive science of science tradition or the jocular boldness with which Derek Price put forward his conjectures. They also tended to be more critical towards the science policy usage of the *SCI*. For example, after Harriet Zuckerman and Jonathan Cole had criticized *SCI*'s abuse "by participants in the decision-making process about promotion and tenure" in *the New York Times*, emphasizing that citation data should not be used for assessing individuals (Charlton 1981), Derek Price reacted strongly:

> I continue to disagree with you about the usefulness of citation data on an individual basis. It is absurd to give someone with 21 citations an edge over someone with 19, but that is simply a matter of calculating appropriate standard deviations and probable errors. We all hope for greater sophistication but the point is that citations do provide a reproducible and clearly useful measure of something. The big question is not whether it correlates with quality in some particular sense of that term but rather what sort of quality is being measured. (...) I believe that even on an individual basis the SCI is one of the very important and useful indices of quality to be correlated with all others that are available and, indeed, with all the other evaluations

[72] Zuckerman to Garfield, July 27, 1971.

[73] Garfield to Zuckerman, August 18, 1971.

[74] This visit took place on November 22, 1971 at the Russell Sage Foundation in New York where Zuckerman was on leave (Memo from Garfield to Malin, December 7, 1971).

[75] Zuckerman to Garfield, October 2, 1972.

> of quality other than scientific research which may go into a decision to fund, promote, or give tenure. We are not at liberty to drag down or badmouth any of these indices. In fact I feel it must be morally unsound to ignore any evidence that is available. I also don't understand what it is you are trying to do by pedaling backward so hard.[76]

4.7 The citation sociologically explained

Notwithstanding these differences in approach, Derek Price and the Columbia sociologists of science had much in common. For one thing, the latter group were immensely influenced by the early works of Price.

> Price was a very important influence on my brother and myself. I mean his book *Little Science, Big Science* was fascinating. I must have read that book over twenty, thirty times. It just led to so many different hypotheses. Let us say that Merton and Price were the two most important influences.[77]

The science of science and the sociology of science both accepted the citation frequency as a valid measure of scientific quality and as a sociometrically interesting link between authors or publications. The premise shared by everyone was that the reference and the citation were basically identical (chapter 1). This enabled an intuitive approach to the meaning of citation. Garfield, Price, Sher, and the sociologists of science at Columbia University all regularly referred to the act of citing to justify using citation data. For the latter group, this was especially appealing because the citation seemed to fit in very nicely with Merton's norms of science (Merton 1973). It was not Merton, however, but Norman Kaplan who provided for the first explicit Mertonian explanation of the citation (Kaplan 1965). From this perspective, the citation is seen as the embodiment of the giving of recognition to which the scientist is obliged. Since this leads to a symmetrical positioning of the citation, it means that, provided the normal statistical precautions have been taken, the number of citations one receives is directly proportional to the recognition one receives. This seemed so obvious, that Merton later wondered why he had not invented a device like the *SCI* himself (Merton 1977):

> Here then[78], with the formidable advantage of hindsight (and its attendant risk of anachronism), we find described all the <u>substantive</u> characteristics of science required for the invention and application of a research tool that is largely specific to the history and sociology of science: the <u>tool of the citation index</u> and the correlative <u>method of citation analysis</u>. But if all the substantive ingredients for invention of that tool were being observed back in 1942, why was the citation index in science not then invented? Why was it clearly described only in the mid-1950s, designed and constructed in pilot form only in the early 1960s and actually introduced as the <u>Science Citation Index</u> only in 1964?

[76]Price to Cole and Zuckerman, May 1, 1981.
[77]Interview with Stephen Cole, November 1994, New Orleans.
[78]Merton refers to Merton (1973), first published in 1942.

The answer Merton gave in 1977 was "simple enough". He characterized his missing the *SCI* as an example of "a total miss": "All of the substantive ingredients for the invention were there — except for the essential ones".

> It was one thing to have this sketchy substantive model of how the reciprocity of roles worked, and quite another thing to have the idea that this composite communication-and-reward system and some aspects of its cognitive output could be investigated by means of a citation index. Absent from this early thinking was the very notion of moving from substance to procedure. Absent was the wit to draw out from the substantive model the several operational implications needed to arrive at the invention. Absent was the specific idea of devising a method for systematically identifying the observable outputs of scientists who were obliged to specify the sources of the knowledge they drew upon, doing so freely and routinely as a result both of having internalized the norm and of having incentive since their work, in turn, might in due course receive the ultimate reward of peer recognition in the same way. (For if one's work is not being noticed and used by others in the system of science, doubts of its value are apt to arise.) Absent was the basic perception of what was so obvious as to remain unnoticed: that there had evolved, long ago, a device for publicly and routinely acknowledging such intellectual debts in the form of the reference and citation. The absence of these basic ideas precluded the further crucial insight that once citations were aggregated, sorted out, and sytematically analyzed, they should in principle reflect cognitive as well as other linkages between scientists, as individuals and as members of latent as well as more visibly organized groups. These were the large and critical conceptual gaps that separated the substantive model from any notion of a citation index as a tool for such systematic, quantitative analysis. (Merton 1977, 54–55)[79]

This partly intuitive, partly theoretically underpinned use of the citation and reference as a device in the reward system of science was the dominant paradigm during the 1960s and the first half of the 1970s.

It was strengthened by the findings of Jonathan and Stephen Cole that the citation frequency correlated statistically significant with several other indicators of scientific quality (Cole & Cole 1971, 28): "The data available indicate that straight citation counts are highly correlated with virtually every refined measure of quality. (...) Consequently, it is possible to use straight counts of citations with reasonable confidence."

This conclusion enabled the seamless combination of quantitative and qualitative methods, an important feature of Mertonian sociology of science. Mertonian theory provided a clear theoretical framework to integrate citation analysis with other empirical methods. Since the decline of functionalism in sociology, and the subsequent advent of constructivism in science studies in the early 1980s, this bond between quantitative and qualitative science studies has been severed. This was not only brought about by changes in theoretical and methodological

[79]Moreover, Merton stressed, there were technological causes, mainly computer technology and information science.

stance within science studies. The separation of the quantitative and the qualitative orientation is also related to the socio-cognitive evolution of scientometrics. While the qualitative science studies have been fascinated by the constructivist paradigm since the early eighties, the quantitative study of science has been captured by the emerging citation culture. Nowadays relatively few students of science combine quantitative and qualitative expertise. This may very well be caused by the seemingly incompatible nature of the science representations they produce. It is not impossible to bridge the gulf between the different representations but it takes more work than it did thirty years ago.

Back to Contents Page

Chapter 5

Back to Contents Page

The signs of science

> we need to know more than is yet known about what references and citations do and do not represent if citation analysis is to provide further understanding of how science is socially and cognitively organized and practised. (Merton 1977)

5.1 Introduction

> **citation, citations** N COUNT; a rather formal word. A **citation** is
>
> 1. an official document or speech which praises a person for having done something brave or special.
> 2. a summons to appear before a court or law; a legal term.
> 3. a quotation from a book or other piece of writing.
>
> **cite, cites, citing, cited** V+O; IF+PREP; a rather formal word.
>
> 1. If you **cite** something,
>   (a) you mention it, especially as an example or as proof of what you are saying.
>   (b) you quote from a written work, especially as an example or a proof of what you are saying.
> 2. To *cite* someone or something in a legal action means to officially mention or name them.
> 3. To *cite* someone means
>   (a) to officially summon them to appear before a court of law.
>   (b) to officially praise them in a report or other document because they have done something brave or special.
>
> (Wil 1991, 243)

This chapter deconstructs a number of exemplary scientometric indicators in order to better understand the nature (or even better, culture) of the citation. The creators of the *SCI*, Eugene Garfield and Irving Sher, with the help of Joshua Lederberg and Gordon Allen, translated the juridical concept of citation indexing into a new one, applicable to science (chapter 2). This translation entailed the creation of a new sign (chapter 1). The resulting new symbolic possibilities stimulated quantitative research within the sociology of science and, furthermore, its exploration according to the science of science tradition (chapter 4). This was only the beginning. The citation has also become the basic building block of aggregate indicators of science. Together they form an intricate maze of signs that provide for a new representation of the sciences. This citation representation is partly, but crucially, influenced by the way the citation is produced. Of course, it is also determined by the distribution of references in the processed literature. To gauge the precise relationships between these two domains it is necessary to probe more deeply into the way the new signs of science have been, and are being, constructed.

5.2 Basic properties of the citation

> All citations are equal. (Smith 1981, 89)

Chapter 1 and chapter 2 showed that the production of citation indexes leads to the transformation of the reference into the corresponding citation. Thereby a new quality is created: “All citations are equal” (Smith 1981). This handy property enables the addition of citations of a given article resulting in the so-called citation frequency. The sign *citation* does not share this feature with its parent, the sign *reference*. References are not equal: they have different functions in the citing text, and their underlying motives are various. It is true, indeed, that the reference can be added too: this gives the number of references of the citing article. But this is its least important feature. The reference functions predominantly as a pointer. With the citation it is the other way around. This sign can function both as a pointer and as a number, but it is the latter function, not the former, that carries the day. Moreover, in as much as the citation points, it does this on behalf of its role as the fundamental unit of the citation frequency. Pointing is, in other words, auxiliary to counting in citation analysis. Not only can the citation frequency of a certain article be measured, this frequency can also be summed up at higher levels of aggregation to obtain the citation frequencies of research groups, institutions, journals, countries and even of disciplines and scholarly fields in their entirety.

Consequently, the citation seems to have a universal quality. Since the citation frequency of every article can be measured — if it is not cited, it can be given a citation frequency of zero — any article can be compared with any other, independently of the subjects involved. Should the citation not have this property, a citation index would be less useful as a bibliographic instrument:

> Another reason why citation indexes are so useful is their independence of topic-descriptors, avoiding the imprecision and inconsistency inherent in the use of such topic-descriptors. (Egghe & Rousseau 1990)

In other words, citations play the role of language: “citations serve as a kind of language system, which can be deployed with greater flexibility than ordinary language" (Small & Griffith 1974). The role of the citation might also be compared with that of money, especially if the evaluative use of scientometrics is taken into account[1]. Whenever the value of an article is expressed in its citation frequency, the citation is the unit of a "currency of science".

The central point here is the universality, created by a double move. First, the local context of the citing document is removed, in the inversion of reference to citation. Then the local context of the indexing institution is deleted, albeit imperfectly. Citation analysis is in this respect like standardization (Latour & Woolgar 1986): "Nothing so marks the creation of universality as the dropping of local subscripts from units that are nevertheless produced in different physical locales" (O'Connell 1993). The citation shares still another property with the signs of money and language: it can only function properly in the midst of other citations. Therefore, citations need to be mass-produced. A lone citation does not make sense. It derives its function mainly from its relations to other citations. In other words, it is self-referential. Whether one tries to map science or to evaluate it, one needs large amounts of citation data. This self-referential property is well-known of indicators like the I.Q. (Woolgar 1991). Luukkonen (1990) signals this as follows: "The basic problem ... is that the 'quality' of work is measured by citations, an indicator to be tested and studied. This implies a circular reasoning: the most highly cited scientists are highly cited because they are highly cited (=good as indicated by the high citation counts)". From this study's point of view, however, this is less a problem than part of the explanation of the power and culture of citation.

As mentioned, a lone citation does not really count. Seen as an isolated sign, it does not possess many qualities. Virtually the only one is its fundamental equality to all other citations. The more interesting qualities arise from the interactions among citations and those between citations and references. A great deal of scientometric and bibliometric research has dealt with both uncovering and realizing the potential patterns in the citation and reference networks in science.

One of these topics is the way in which "time" is present in the citation representation of science. Again and again, *SCI*'s inventor Eugene Garfield has stressed an important advantage of the citation index as a search tool: if one takes a publication one can track its "descendents" (the articles which cite the publication at hand) up to the present. Whereas snowballing by using references only leads one further and further into the past (because with this method one can only track its "ancestors"—the articles which have been cited by the publication at hand), a citation index brings one closer to the present. The reason for this is obvious: a citing text is necessarily more recent than the related cited one[2]. Because of this, the citation frequency of an article is a dynamic property. It may change at any moment. This raises two points. First, citation networks are always drawn with hindsight; they are by definition a posteriori. Second, existing

[1] At the Technical University of Ankara, the relationship between the citation frequency and money is even more direct. Researchers there earn money by receiving citations. When one has collected a specific number of citations, one is even entitled to become a professor (Prof. Ali Azun, Ankara Technical University, personal communication June 1997, Jerusalem).

[2] The only exception to this rule are references to forthcoming publications.

literature that gives rise to citations is implicitly treated as a static whole. Any article, no matter how old, can in principle be cited. Time is visible in the citation network, not as something that flows nor as movement of the network, but as the absence of certain citation relations (more recent articles not getting citations from older ones), in other words as structure. This is the consequence of the constraints the reference imposes on the citation.

From the early days of the *SCI* citations and references have been turned into composite indicators. Many of these indicators are supposed to capture a specific phenomenon in science. They represent the reality of science. *Raw citation counts* dominated scientometrics in the early years. In the course of the seventies and eighties, they gave way to more refined science indicators. For example, the chance of being cited varies greatly per scientific specialty, dependending on its size and specific citing culture. To deal with this, several types of *normalized citation frequencies* were constructed. Citation counts were created at various levels of aggregation, from the individual researcher to countries. To gauge the citation frequencies of scientific journals, the so-called *Impact Factor* was developed. It is seen as a measure of a journal's impact on subsequent scientific work. Inspired by the concept of bibliographic coupling in the field of library science, the *co-citation frequency* was almost simultaneously and independently created by the American Henry Small (Small 1973) and the Russian Irina Marshakova (Marshakova 1973). This indicator records the number of times two publications are cited together, and is taken to be a measure of similarity of the two publications. Because one can measure co-citations at several levels of aggregation, the co-citation frequency can be used as a building block of scientific cartography: the mapping of science on the basis of co-citation links between publications (Starchild et al. 1981, Garfield et al. 1984). Before these more complex indicators can be deconstructed, however, it is useful to take a closer look at the symbolic processes underlying the production of the citation itself.

5.3 Producing citations

> It seems to me a great pity to waste a good technical term by using the words citation and reference interchangeably. I, therefore, propose and adopt the convention that if Paper R contains a bibliographic footnote using and describing Paper C, then R contains a reference to C, and C has a citation from R. (Price 1970)

Chapter 1 argued that the signs reference and citation should be distinguished from each other. The latter results from the former. This means that the semiosis of the citation is a second order operation with respect to creation of the reference by the scientist. Various interventions in this production process will therefore influence the outcome. First of all, it is impossible to process every reference made in every scientific article in the world. Hence, a selection must be made, which of course influences the resulting index. The way in which the index represents the literature is also modified by the way in which the reference represents the cited text. For most purposes, it really does matter whether or not the citation is a "correct" inversion of the reference: the semiosis of the citation is a precision operation. The actors involved, whether scientometricians or indexers, usually relate to this in terms of the identity of reference and citation: they should be the same. From the point of view developed in this study, it is not a matter of identity but of "true inversion", meaning an inversion without any other changes. No operation can proceed perfectly[3], and this is also true of citation indexing. Therefore, the indexes inevitably contain numerous "mistakes". How serious these are depends on the use of the index. Any citation analysis that processes large amounts of citation data will end up with considerable numbers of these "errors", which cannot easily be identified as all citations look alike. As a consequence there is continuous discussion on the "quality of the data" in the field of scientometrics. In fact, three different topics have been discussed under this one heading: the quality of the reference (section 5.3.1, page 111); the selection of the reference (section 5.3.2, page 112); and the integrity of the inversion (section 5.3.3, page 114).

5.3.1 The quality of the reference

It is possible for the scientist to refer to nonexistent texts, simply by typing the wrong page number or year or making a spelling mistake. Whenever this happens the citation indexer may create a citation as an attribute of a non-existing text. The act of making the citation is reflexive towards the giving of reference. Therefore, it is possible to refer back to the reference in order to detect or correct some of these mistakes. For example, Moed and Vriens (1989) found recurring problems by comparing different databases:

> Roughly speaking, for every ten citations containing a bibliographic description identical to that of a particular article in our target dataset, one citation shows some kind of discrepancy. (...) the major part of the discrepancies in our dataset are due to errors or variations in cited references that are present in the original text.

These authors attribute this deficiency in the references to certain habits of citing authors: "we found evidence that copying references from other articles may be a cause of the observed multiplication of errors in cited references". Another, major, problem is the spelling of journal names. At ISI, journal titles were often found to be abbreviated differently: "there were more than 100,000 different abbreviations for the 12,000 individual journal titles cited in the 3-month sample. Inconsistency was made worse by inaccuracy" (Garfield 1970). Correction of these inconsistencies and inaccuracies has even become common practice wherever they can be traced by comparing citations, for example where journal titles show discrepancies:

> ISI's data unification process can result in a more accurate presentation of the citation data than presented in the original journal literature. (SCI Guidelines for Interpretation of ISI Citation Data)

[3]With the possible exception of digital copying.

This underlines the recursive character of the production of citations. If the reference were the citation, ISI's correction procedure would be sheer magic. It is precisely because of the difference between them, that comparison and correction is possible.

5.3.2 The selection of reference

Selecting references entails choosing which type of textual references to process as references, as well as deciding which journals to use as sources. With the exception of the *Arts & Humanities Citation Index*, which also processes implicit references, only clearly identifiable footnotes and endnotes have been used. This practice has been criticized by Hicks & Potter (1991) who, as only few authors have done, have paid special attention to "the way a citation is produced as a separable, countable category" (Hicks & Potter 1991). In their opinion, the *SCI* does not give a fair representation of the texts by which scientists have been influenced. It only makes use of a thin slice of potential references. Since every text should be seen as "suspended in the network of all others", any part of any text is some kind of unacknowledged reference. These implicit references are not utilized by the *SCI*. At the same time, the most recent, informal, influences on scientists are not transformed into footnotes at all, for the very reason they are informal. Therefore:

> Citations form just a thin but glistening band, sandwiched between the rock of eons. And it is this highly limited, highly unrepresentative, yet alluringly available band of rock that the ISI has fetishized and turned into a highly desirable and marketable commodity. (Hicks & Potter 1991)

Hicks & Potter (1991) have still identified the reference with the citation. Seen from the perspective of this study, their geological metaphore needs to be interpreted in a somewhat different way. Actually, these authors are not so much discussing the geology of citation itself as the geology of its raw materials, the references. ISI's digging up of only a part of the potential reference source is not a coincidence but a consequence of the constraints imposed by the economics of the production of citations. Hicks & Potter's (1991) metaphor is now especially relevant, digging up the other types of references would simply be too expensive:

> The production of citation indexes is more involved than is generally appreciated. Although citation indexing eliminates the expensive intellectual effort associated with traditional subject-term indexing ... producing a citation index of appreciable size is a massive materials-handling and information-processing job. (Garfield 1979)

Therefore, to be converted into a scientometric citation, a reference must exist in an easily recognizable format. The mass production of citations depends on standardization. Not coincidentally, Garfield, Price and their colleagues frequently campaigned for changes in referencing policies of scientific journals. They tried for example to convince editors of the increasing importance of the reference format: "Now that citation indexing has become a valuable and integral part of the computerized systems by which we gain access both to archival and to research-front literature, the editorial practices regarding citations may need re-examination" (de Solla Price 1969). This was not only in the interest of the researchers but of the indexers as well. The more standardized the reference, the less costly the production of the citation.

The second aspect of the selection of raw materials for the mass production of citations is the choice of source journals. This has been given considerable attention by the creators of the *SCI*. It should be remembered, however, that they were building a bibliographic tool, i.e. a search instrument. To function as a search tool for science as a whole, the *SCI* should be able to provide access to scientific literature in its entirety. This does not mean, however, that it should contain a representative sample of the literature, in fact the reverse is true. The distribution of characteristics is often skewed in bibliometric and scientometric research. According to "Bradford's law" (Bradford 1953), the distribution of scientific literature in a certain domain spreads out over many neighboring fields. Garfield (1979) compares this phenomenon with a comet having a concentrated core as well as a widely fanned out tail. "Garfield's law" expands on this by stating that the tails of disciplines show a very significant degree of overlap. This means that there should be a core of all scientific literature, which receives the majority of references. It is Garfield's explanation of why the science citation index is technically feasible in the first place.

> So large is the overlap between disciplines, in fact, that the core literature for all scientific disciplines involves a group of no more than 1000 journals, and may involve as few as 500. In less abstract terms, this means that a good general science library that covers the core literature of all disciplines need not have any more journals than a good special library that covers al the literature of a single discipline. (Garfield 1979)

Griffith et al. (1977) report on the same issue:

> The quality and quantity of the scientific literature 'channelizes'. That is, a combination of social and probabilistic mechanisms ensure that most documents of a discipline, and nearly all documents of the highest quality, appear in a limited number of resources (i.e., journals in the natural sciences). Furthermore, all such important sources may be readily recognized and ranked along this quality dimension by citation counts. (Griffith et al. 1977).

Interestingly, this ranking is used to gauge the selection of source journals:

> The selection of journals is crucial to the success of a citation index because it is a strategy quite different from the usual librarian's striving for completeness. ... the ultimate test is provided as feedback from the journals which are cited by such sources. For many years the list of cited journals has provided a higher criticism of which journals to accept and which to reject as sources. (...) Thus although it [the *SCI*] is derived from only $\frac{1}{15}$ of the source papers, it includes $\frac{3}{4}$ of the cited literature. (Price 1979)

Not surprisingly, the citation quickly evolved into an indicator of the appropriateness of the selection of source journals in the production of citation indexes: "The list of most frequently cited journals shows that the SCI has been remarkably successful in covering all 'significant' and 'important' journals, insofar as citation counts can be considered a reliable measure of 'importance' and 'significance" (Garfield 1970). This relates to the self-referential property of the citation index. Often, the use of citation indicators can only be justified in terms of these indicators themselves. For example, the question of whether the *Science Citation Index* uses the appropriate source journals has been answered by measuring the citation frequency of these source journals. This is a different criterion from the representativity of the source journals, if only because ISI selects the highly cited journals, while most journals are, by definition, less cited. In fact, if the citation index is to function well as a bibliographic search tool, its selection of source journals should nót be a representative sample of science as a whole.

5.3.3 The integrity of the inversion

The third topic is the inversion process itself. This process may create additional problems, which didn't exist with the original reference. Again, the import of irregularities depends on the later use of the citation. Nevertheless, scientometricians have repeatedly compiled long lists of problems encountered while using ISI's data. Most of these arise because of the inherent uncertainty of the inversion. Smith (1981) gives an overview of the difficulties that may be created by the production process of the citation. They vary from ISI's policy of registering only the first authors of cited texts, and the regular occurrence of identical names for different entities, to field-dependent differences and plain errors[4]. Egghe & Rousseau (1990) give some additional problems, like the incompleteness of the ISI database, the dominance of English as a scientific language, and the American and sex bias. Cleaning up the data sets is common practice at scientometric indicator and research centres like the Centre for Science and Technology Studies in Leiden, CHI (the company led by patent analysis pioneer Francis Narin) in New Jersey and the Information and Scientometrics Research Unit of the Library of the Hungarian Academy of Sciences in Budapest. Given the labour intensive character of recursively improving the quality of the data, these databases are jealously guarded by their proprietors: access to clean data is a strategic opportunity in scientometrics.

[4]Errors may have quite unpredictable consequences because of the skewed distributions of bibliometric data: "citation data are extremely positively skwewed. (...) This leads us to a second point. The presence of errors is a nontrivial factor. (...) a random error rate of nearly 20 % was probably present in our study (...) How does the 'signal' retained vary with the type of citation count performed? If one loses 33% of an individual's citations (co-authored ones), fails to cull 5% co-author self-citations, and adds a substantial random error rate, what does one have? Automated search may surrender 25% of the sample. Further, estimates of the proportion of 'perfunctory' citations in high-energy physics range from about 20 per cent to 40 per cent. The value of a citation count is thus a complex function of the type of counting and the intended uses; it cannot be taken at simple face value" (Porter 1977).

5.4 Building upon the citation

Raw citation counts have given rise to more sophisticated indicators. They are too numerous to discuss individually, and neither would it be particularly enlightening. There is, however, a pattern in the construction of aggregate indicators. First, all are built on the basis of varying combinations of the signs *reference* and *citation*. The way these two dimensions (the citing and the cited) are combined determines important characteristics of the resulting indicator network. Second, they all aim to represent reality in a more reliable way than competing indicators, or qualitative descriptions. Third, they build upon one another.

5.4.1 The Price Index

The Price Index is the simplest aggregate scientometric indicator.

$$PI = N_1/N_2 \times 100 \tag{5.1}$$

where N_1 is the number of references which were published less than six years before the citing publication, and N_2 is the total number of references.

In other words, as it represents the number of references to the last (five) years as a percentage of the total number of references, the Price Index is a measure of the recency of the literature cited by a given article, journal or specialty. The Price Index can be computed for a given year (Price 1970), or alternatively on a per article basis (Moed 1989). The Price Index is a pure reference indicator. It is indicative of citing behaviour. Whereas Derek de Solla Price took this measure as a rather straightforward means of distinguishing the hard sciences from "the soft ones and from non-science" (Price 1970), more recent research has shown this to be an oversimplification (Moed 1989). Although the index itself is very simple, its sociological interpretation is not without ambiguities.

5.4.2 The Impact Factor

Garfield's Impact Factor is an example of a purely citation based indicator.

$$IF = C/N \tag{5.2}$$

where

- C is the number of citations a scientific journal begets in a certain period,
- N is the number of publications in that journal during the same period.

In most cases a period of two years is standard as the basis for the computation[5]. This is for example the basis of the impact factors published in ISI's *Journal*

[5] This seems to have been decided, rather casually, on the basis of a general impression of the first raw data on the variation of citation frequencies with the age distribution of cited articles (Interview with Irv Sher, Philadelphia, 1992).

Citation Reports. Of course, different choices of time periods give rise to different values of the Impact Factor. This is the main reason that, simple as the indicator may seem, the application and construction of the Impact Factor has been a point of contention in scientometrics (Moed & van Leeuwen 1995). The Impact Factor was created to be able to compare scientific journals with respect to the probability of their being cited, and is an example of normalization:

> Citation frequency is, of course, a function of many variables besides scientific merit. ... Citation frequency of a journal is thus a function not only of the scientific significance of the material it publishes ... but also of the amount of material it publishes. (Garfield 1955)

A journal's Impact Factor represents the probability of being cited if one is published in that journal. Therefore, this index creates the citation frequency as a property of the journal. The division by the number of citeable items enables one to compare journals with widely differing publication frequencies or number of articles.

5.4.3 Co-citation clustering

Co-citation analysis (Small 1973, Marshakova 1973) was the principal instrument in ISI's Atlas of Science project (Starchild et al. 1981, Garfield et al. 1984), and is widely used by scientometricians (Some of the early studies are discussed in Small & Sweeney 1985*b*, Small & Sweeney 1985*a*, Small & Griffith 1974, Small 1977, Hicks 1987, Hicks 1988, Franklin 1988, Sullivan, White & Barboni 1977, Edge 1977, Oberski 1988, Oberski 1987). Its basic entity, the co-citation frequency, is the number of times a certain pair of cited articles are cited together.

$$CCF_{i,j} = C_i \cup C_j \tag{5.3}$$

where C_i is the set of citations to article i.

Basically, the co-citation frequency is the common occurrence of two references in a bibliography of a citing article. In this sense, the co-citation frequency is built upon the reference. As Egghe & Rousseau (1990) show, the co-citation frequency can also be computed by taking both sets of citations of the two cited articles and measuring the intersection of these two sets. In this case, the co-citation frequency is the number of citations in this intersection. Like the citation frequency, the co-citation frequency is a number. It is supposed to be a measure as well as proof of the existence of a symmetrical relationship between two cited documents.

The whole point of co-citation clustering is its capacity to create maps of science that can be interpreted by scientists, science managers or science policy officals. This does not mean that co-citation clustering straightforwardly reflects the true nature of science. The reality of co-citation clusters is, on the contrary, the very consequence of built-in inconsistencies and "ontological gerrymandering" (Woolgar 1991). In this respect, co-citation clustering tells an important part of the story of scientometrics. It is worthwhile taking a closer look at what precisely happens in co-citation analysis.

The co-citation procedure starts with the selection of highly cited articles (Small & Griffith (1974) mention a threshold of 10 citations). This list of highly cited articles is then inverted to retrieve the articles that cited them. This results in a selection of lists of highly cited references and their citing articles from all references processed by the citation index. Then the number of times each pair of these references co-occurs in a bibliography is counted, and the whole file is resorted. This gives a long list of pairs of cited articles and their co-citation frequency. The citing articles are subsequently discarded. This transforms the co-citation frequency from the number of times two independent, asymmetrical citing relations appear together, into a measure of a new symmetrical relationship between the two cited documents. With this inversion, we have entered a new domain, the world of co-citation.

The next step is a clustering procedure. Small & Griffith (1974) use the so-called "single link clustering" in which one co-citation link between documents is enough to include that document in one cluster:

> Typically, a cluster would be generated by selecting a starting document and listing all other documents with which is was paired. These new documents were added to the cluster, and the documents they were with were, in turn, added to the cluster. This process was continued until all documents which were linked with those already in the cluster had been identified. (Small & Griffith 1974)

Thresholds are used in this clustering procedure to distinguish different levels of co-citation relationships. At the lowest level, all clusters are connected. Setting the threshold higher, gives the effect of "zooming in" on a certain cluster. The rest is a matter of display. Lists of clusters can be generated; alternatively, multi-dimensional scaling techniques can be used to generate maps. Recently, ISI even developed software to enable the drawing of these maps with desktop computers.

Thus, several translation steps are involved in co-citation clustering, the inversion of the co-occurrence of two references into the co-citation frequency being the crucial one. It should be stressed that this inversion is essentially the same one we have already seen in citation analysis. Again, a consequence of citing behaviour is transformed into a property of cited documents. Depending on the algorithm used, this inversion can be performed as an early or a late step in the computations. If Small & Griffith's (1974) procedure is used, inversion is one of the later steps. If we compute the co-citation frequency according to Egghe & Rousseau (1990), we start with the inversion. The end result is the same.

The science maps based on this co-citation clustering have been seen as the "true reflection" of the structure of science. Small & Griffith (1974) report that the clusters they found consist of:

> a relatively small group of co-cited documents and a larger group of documents, each of which cites two or more of the co-cited documents. These two sets of documents are believed to form the basis of a research area, a group

> of researchers engaged in the joint exploration of related problems. (Small & Griffith 1974)

This interpretation is based on the assumption that the existence of a co-citation link indicates some kind of transfer of information or cognitive relatedness: "We will assume that the larger the number of co-citing documents, the greater the amount of transfer and exchange of information within and between specialties". Small & Griffith (1974) found a large number of biomedical documents, forming one overwhelming cluster. It was interpreted in a realist way: "These findings reflect, on the one hand, the strong representation of the biomedical literature in the SCI and, on the other, the pervasive use of certain standard methods and procedures in biomedical work". Subsequently, they broke up this big cluster into 65 smaller ones by raising the threshold and, at the same time, removed the linking method papers before reclustering: "It appears, from these data, that the biomedical literature presents a structure that is quite different from the physical sciences; and, secondly, that some comparatively simple strategies can be employed to break up clusters which are pulled together by papers whose use is not confined to a single specialty".

The "ontological gerrymandering" of co-citation analysis (and citation analysis at large) is clearly demonstrated in the reflexive improvement of the technique developed in 1985. What was seen by Small & Griffith (1974) as the reflection of the reality of science, is perceived by Small & Sweeney (1985*b*) and Small & Sweeney (1985*a*) as a methodological problem to be solved by changing the method of co-citation clustering. Far from being the result of science itself, the large biomedical cluster appears to be an artefact of the clustering method: "it proved difficult to obtain an adequate representation for fields such as mathematics and engineering within the broad mix of biological and physical sciences. ... These facts pointed strongly to a biomedical over-representation in the annual SCI clusters". This is caused by the higher referencing level in biomedical literature: "in two ways: 1: by increasing the number and proportion of biomedical items which fall in the highly cited range ... and 2: by increasing the strength and intensity of co-citation links formed among biomedical items". After the methodological improvement, the goal of clustering is stated as follows:

> A reasonable objective seems to be that the number of clusters for a field be proportional to its source article representation in the data base. If five percent of the articles in a year are in mathematics, then five percent of the clusters should be on mathematical topics. The problem becomes how to compensate for the differences in referencing patterns from field to field and indeed from article to article.

Interestingly enough, the epistemological status of ISI's journal selection is fundamentally altered in this argument. While Garfield and his colleagues never pretended to draw a representative sample of science, the proportionate composition of the source journals of ISI is used by Small & Sweeney (1985*b*) and Small & Sweeney (1985*a*) as a baseline for the validation of the clustering technique in precisely such a way.

Small & Sweeney (1985*b*) and Small & Sweeney (1985*a*) reach their goal in two steps. In the first place, they introduce fractional citation counting to get a "better" representation of the disciplinary structure of science. This is a subtle but important innovation: no longer does every citation count as being equal. Instead, every citing article is assigned a value of 1 to be evenly distributed over its references. Every reference, therefore, gets the value of the inverse of the number of references.

> The important concept here is that all source items have but a single unit of credit to dispense. (Small & Sweeney 1985*b*, Small & Sweeney 1985*a*)

The reference transfers this value to its corresponding citation. The citation frequency is then the sum of these different fractions. The citation frequency is based on: the equality of the citing articles; the relative equality of the references of a given citing article; and the possibility of adding the resulting values of the citations. Where the propensity of the citation to be added was based on the citation itself as the fundamental unit (with a value of 1), fractional citation counting takes as its fundamental unit the citing article (which can distribute a value of 1 over its references). The co-citation frequency is thus a hybrid entity composed of the citing and the cited context. Since it relates the number of citations of cited articles to the number of references per citing article, it diminishes the domination of highly citing fields, like biomedicine, in co-citation based science maps.

The second innovation introduced in co-citation clustering is the replacement of fixed thresholds as cluster criteria by variable ones. Small & Sweeney (1985*b*) and Small & Sweeney (1985*a*) note that the single-link clustering algorithm, with its low demand of only one link, leads to "chaining", the creation of large macro-clusters. Whereas Small & Griffith (1974) saw this as proof of the interconnectedness of science, Small & Sweeney (1985*b*) and Small & Sweeney (1985*a*) speak of it as a technical problem in need of repair by a new cluster-defining technique. Instead of a fixed threshold, a fixed upper value of the cluster size is defined with a reference to Price's invisible colleges (Price 1965*a*). Whenever this cluster size is exceeded, the clustering procedure starts anew with a higher threshold. This has the advantage that it "also prevents the formation of amorphous macro-clusters by chaining, which is a problem with the single-link method when low co-citation levels are used".

In the selection of highly cited items, fractional citation counting is used, but integral counts are the basis of the co-citation frequency itself. The authors, wishing to use the two methods independently, even stress this point: "it is important to note that we did not use the fractional counts for the normalization of integer co-citation counts. The fractional counts were used only in the initial selection of cited items".

It is therefore possible to go one step further and apply an analogous fractioning procedure to co-citation counts. And this is indeed what the authors speculate on:

> fractional co-citation would assign a single unit of co-citing strength to each citing paper among all the pairs of references cited by that paper. If, for

> example, a paper cited "n" highly cited items, each pair of cited items would be assigned a weighted co-citation equal to $\frac{1}{2n\times(n-1)}$. The summation of all such fractional cocitation contributions from all citing papers for a given pair of cited items would constitute the fractional co-citation count for that cited pair. (Small & Sweeney 1985*b*, Small & Sweeney 1985*a*)

The authors expect that using this would further extend the "balanced" representation of science, "to structural features of the fields as well, e.g. the density of links". With these fractional co-citation counts, the fractioning is only over the number of highly cited references, not over all of them. Its resulting fractional co-citation frequency is based neither on the equality of all citations (as the integral co-citation and citation frequency are) nor on the equality of the citing articles (as the fractional citation frequency is). It is founded on: the relative equality of all articles that co-cite at least one pair of highly cited articles; the relative equality of the co-occurences of highly cited references in a given bibliography; and the possibility to add the resulting fractional co-citation counts. This last capacity is neither based on the citation nor on the citing article, but on the article-that-co-cites-highly-cited-articles as the fundamental unit with a value of 1. This further increases the relative inequalities of citations.

In summary, co-citation analysis liberally juxtaposes elementary references and citations to build an intricate network of complex indicators and relations. Its central feature is the inversion of the co-occurence of pairs of references to the co-citation of pairs of cited documents. This inversion creates additional degrees of freedom of the sign cocitation frequency and gives the technique the flexibility to develop on the basis of its own results and shortcomings. This leads to the ingenious construction of ever more elaborate and abstract indicators, in which the citing and the cited dimensions of scientific literature are freely mixed. In no way is there a fixed boundary between the hard reality of scientific literature and its representation in co-citation clusters. On the contrary, this boundary is the result of the application and interpretation of the technique.

The method seems to develop itself in two ways. First by extension through analogy, as we have seen with the expanding domain of fractional citation counting. Second through deconstruction of parts of the foundations of co-citation analysis and reconstruction of new ones without the method becoming invalid in the eyes of its proponents. In other words, it is not the consistency of the method that gives it its power (Ziman 1979) but the very lack of it. Co-citation analysis flourishes on its contradictions. What was an interesting part of the reality of science in 1974, was an artefact of deficiencies of the technique in 1985. This improvement resulted in a reconstruction of the boundary between reality and representation. Of course, this process is infinite.

This improved co-citation clustering has indeed been deconstructed by another indicator building group. First these authors reiterate the existing state of affairs from their point of view:

> Main problems in co-citation analysis concern the occurrence of artefacts (clusters are mainly the result of the applied technique), the stability of cluster structure (continuity over time) and the interpretation of the results. ...

> Both decisions (setting threshold levels for citations and for co-citations) may influence the clustering results, and the rather arbitrary way in which these decisions are made, leads to severe problems of interpretation and evaluation of the results. (van Raan 1988)

The authors wonder whether the cluster structure created by the existing co-citation methods is indeed the most appropriate one when compared to other possible cluster structures:

> "What criteria can be used to specify thresholds in a less arbitrary way? A choice can be made experimentally by looking at the interpretability of the resulting cluster structures, or by applying some a priori criteria. At the present moment there is no definite method for either of these approaches.

These authors find that, contrary to Small's experiences, journals do not discriminate very well between clusters, whereas indexing terms do. They conclude that a much lower threshold than normally applied should be used; they did not find macroclusters with low thresholds, "at least as long as it is not very low". Getting relatively stable results, the authors are able to reconstruct a new boundary between their results and their method: "This means that, at least in our case, the cluster structure is not a simple artefact of the technique".

Part and parcel of this reconstruction of the difference between the reality of science and the artefacts of the method of co-citation analysis is the subsequent validation of the results. From the very beginning of co-citation analysis the clusters have been related to the perception of the citing scientist. This is no coincidence, for the origin of the co-citation frequency is the co-occurrence of two references in a bibliography, put there by an author of a scientific article. Small & Griffith (1974) saw the appearance of clusters as proof of the existence of specialties or invisible colleges. More precisely, they thought that co-citation clustering recreated the structure of science as scientists themselves perceive this structure:

> Assuming that highly-cited items reflect the significant concepts in a field, then co-citation associations between them represent clusters of related concepts. (...) Using the cluster maps in this volume, one can, at a glance, trace the historical developments of an area of research, and identify the papers that made the most important contribution to its growth. (Starchild et al. 1981)

At the same time, co-citation clusters are supposedly robust, thanks to the selection of highly cited documents from the wealth of scarcely cited texts as starting point of the technique: "The idiosyncratic citation behavior of some scientific authors is likely to have little effect upon the patterns being observed here". This robustness has been underlined by validation of the co-citation clusters, which has been done in various ways. Again, these methods exemplify important traits of scientometrics at large. The most direct way is noting, as Braam et al. (1988) have done, that earlier problems have disappeared, e.g. the macro-clusters that plagued Small and his colleagues. Secondly, the resulting clusters can be compared with other, independent criteria, like the title words in the citing or in the cited articles of the cluster (Small & Griffith 1974). Thirdly, scientific experts in the fields involved may be asked to take a look at the maps of their specialties and give their own post-hoc interpretation. Of course, these validation methods frequently give only partial validation. In its turn this may give rise to a new round of partial deconstruction of the technique in order to get rid of the pertinent problems. The essentially partial and contradictory nature of these validation procedures is illustrated by Sullivan et al. (1977). In writing the history of a specialty, they found that co-citational history and plain historical intellectual history

> did not produce inconsistent results, but they did convey some different kinds of information. In particular, the cocitation analyses provided dramatic indicators of shifts in the dominant foci of intellectual activity that were harder to see in the data we used to write our intellectual history.(Sullivan et al. 1977)

Combined, the two methods proved "very useful": "Cluster formation, in other words, signals the broad acceptance of a line of research, not necessarily its actual inception ... The assumption that articles which define clusters share in the research tradition of the clusters would therefore seem justified". However, articles were missed by co-citation because they cited only one document in a cluster, or they cited none of the articles.

> Ironically, the more important, salient, and accepted the problem focus, perhaps the less accurate the retrieval of papers via co-citation.

The articles retrieved were, according to these authors, neither representative, nor a constant fraction of the literature of the specialty. Therefore, co-citation analysis was an unreliable measure of the growth of a specialty, because the thresholds are held constant, while the focus of the intellectual activity influences both the size of the clusters and the fraction of the total population. The claim that "the mechanical production of co-citation analyses of scientific literature will lay bare the structure of science" is in the view of Sullivan et al. (1977) "too strong".

A well-known critique of co-citation analysis has been written by Edge (1977). "If", argues Edge (1977), "co-citation analysis gives an objective account of science,

> then why bother to 'validate' co-citation studies? Differences between the co-citation results and those derived from other sources are only to be expected, and it is implicit in the method that preference, in such cases, should be given to the former over the latter. However, co-citation practitioners lose nerve at this point: not only do they undertake validations, but they allow *errors*.

Interestingly, Edge's critique is based on the same assumptions with respect to the citation as co-citation analysis itself, especially his supposition of consistency. Edge tries to keep co-citationers to their own words, criticizing them for not being courageous enough. What is seen by the co-citation analists as an argument for the reality of their clusters, is seen by Edge (1977) as a reason to strongly oppose this type of research.

This deconstruction of co-citation analysis confirms the conclusion that the validation procedures contribute to the reality of co-citation clusters, but for different reasons. The construction of co-citation indicators is neither simple nor self-evident. Validation has been shown to occur not to a robust reality "out there". Rather, this is done by comparing one representation of science with another, often an expert's. This results in a mixture of difference and sameness. By stressing either one or the other, different realities are constructed by juxtaposing these different representations. It also enables the incorporation of criticism in new techniques and methods, at the same time defying the ultimate grounding of co-citation analysis on some definite truth as well as evading a kind of definite critique that mistakes scientometric ideology for its practice.

The difference between different representations is illustrated by the representation of time in co-citation analysis. As noted, time is represented in the citation network by missing citation relations. Co-citation analysis, if applied in writing the history of a specialty, must be carried out for several consecutive years. However, there is no given continuity between these years, contrary to more traditional ways of writing the history of science. After all, every co-citation map is drawn with hindsight from a given citing year to all previous, potentially cited, literature. Even if the clusters were to directly represent the perception of scientists, the writing of a co-citational history would be like a flickering series of flashbacks in a movie without there necessarily being any connection in between. Although co-citation analysis is justified as giving the collective perception of the active research community, it is at the same time supposed to enable the continuous history of science. In the first statement the origin of the co-citation frequency is stressed, whereas in the latter the robustness of the resulting clusters is emphasized. Co-citation analysis results in successive slices of frozen science literature, each slice being reconstructed with hindsight from a given year. In a way, co-citation analysis therefore results in a successive, but discontinuous, series of histories of science. No wonder clusters are prone to sudden changes (Sullivan et al. 1977) that may or may not be found in other historical representations.

5.4.4 Normalization procedures

Normalization is a very common procedure in citation analysis and scientometrics. Several ways of normalizing citation counts have already been shown in this study. Garfield's Impact Factor is one example, normalizing the citation counts to the number of publications; Small's fractional citation counting is another one, relating it to the number of references of citing articles. A different type of normalization is used to relate the citation to itself. Then, the raw citation counts are divided by some weighted average number of citations in a certain domain. These methods stick strictly to the perspective of the cited article and see citations something like raindrops falling on articles. In environments where it often rains citations, articles tend to have a higher citation frequency than in other environments. Hence, if one wishes to distinguish articles on an individual basis, it makes sense to construct relative citation counts by dividing the raw number of citations by the number of background citations. The resulting indicators are especially relevant to evaluations of scientific articles, journals, institutions and scientists. As has been shown, they also play a minor role in co-citation analysis: if the analyst wishes to normalize the co-citation counts to the number of citations of each cited article, the so-called Jaccard coefficient — the co-citation counts of a pair of cited articles divided by the root of the product of the citation counts of each article — is one of the possible indicators.

$$JC_{i,j} = \frac{CCF_{i,j}}{\sqrt{C_i \times C_j}} = \frac{C_i \cup C_j}{\sqrt{C_i \times C_j}} \tag{5.4}$$

In other words, the Jaccard coefficient shows the co-citation frequency, the intersection of the two sets of citations, as related to the two sets. Co-citation is then seen as contingent upon citation, the last providing the background for the first.

The citation frequency is, however, not the only possible frame of reference for normalization. This procedure may also be applied to other dimensions in science. The most frequently used are the number of citing publications, the number of co-authors of the cited articles, and the number of references of the citing texts. The normalization of the citation frequency to the number of co-authors is especially relevant if authors are to be evaluated on the basis of the citation frequency of their work, or if these citation counts are to be used at a higher level of aggregation like in comparisons between countries. In both cases the question arises of how to divide the citation counts over multiple authors of cited articles. Within scientometrics, three solutions of this problem have been devised (Egghe & Rousseau 1990): counting only the first author; giving each author the full credit; and dividing the credit over the authors. The first method is based on the fact that the *SCI* used to list only the first authors. The second method is based on the notion that all citations are equal, as well as all authors, and transfers the whole citation frequency from the cited article to every author. The third and last one is based on the equality of all co-authoring authors in dividing the citation frequency of their article among themselves. Consequently, citations are not identical anymore; they are less valuable if they refer to articles with multiple authors. This is an example of fractional counting from the cited perspective, whereas Small's fractional counting is an example of fractioning from the citing perspective.

The coupling of the cited and the citing dimension plays an important role in the methodology developed in the Science Indicators Project of the United States National Science Foundation. In this policy context, the Influence Methodology has been developed by the group of the physicist Francis Narin (Narin 1976). This method starts with the construction of a citation matrix in which the value of cell_j^i is determined by the number of references i gives to j.

$$\begin{Vmatrix} C_1^1 & C_2^1 & .. & C_j^1 \\ C_1^2 & C_2^2 & .. & C_j^2 \\ .. & .. & .. & .. \\ C_1^i & C_2^i & .. & C_j^i \end{Vmatrix} \tag{5.5}$$

where C_j^i is the number of references unit i gives to unit j. The units can be any "publishing entity", like articles, authors, journals or specialties. Narin (1976)'s method is based on the following assumption:

> The citation matrix is the fundamental unit which contains the information describing the flow of influence among units. ... The citation matrix may be thought of as an 'input output' matrix with the medium of exchange being the citation. Each unit gives out references and receives citations; it is above average if it has a 'positive citation balance', i.e. receives more than it gives out. (Narin 1976)

In this matrix reference and citation are identical, signalling the transfer of a fundamental unit of "influence". Every unit in the matrix starts with an individual weight that is obtained by dividing the number of citations it gets from all other units in the matrix by the number of references it gives to all other units.

$$W_j^i = \frac{\sum_1^k C_i^k}{\sum_1^k C_k^i} \tag{5.6}$$

This is only the beginning, however. The central notion of the "Influence Methodology" is the relative character of influence. The weight of each unit (its influence balance) is supposed to determine the weight of the references it gives to other units. In other words, each citing unit conveyes its influence weight to every reference, which again transfers it to the cited units. In this second step, the initial equality of all citations and references is abolished and replaced by a weighted set of values, determined by the initial distribution of citations and references. As a consequence, the weight of every unit is dependent on the weights of all other units and must be determined in an iterative procedure of many more steps. The influence weight of a unit is defined as the product of the citations to the unit and the relative weight of the citing units, summed over all citing units, and divided by all references from the unit.

$$WW_j^i = \frac{\sum_1^k (C_i^k \times W_i^k)}{\sum_1^k C_k^i} \tag{5.7}$$

> The n equations, one for each unit, provide a self consistent 'bootstrap' set of relations in which each unit plays a role in determining the weight of every other unit. ... This procedure is closely related to the standard method for finding the dominant eigenvalue of a matrix. (Narin 1976)

On the basis of this network indicator, hierarchies of journals and maps of specialties have been constructed by ordering according to influence. Narin's influence measure is related to Small's fractional citation counting, which is the first order approximation of Narin's measure if articles are the units in the citation matrix. The difference is in the goal of the method. Whereas Small only wished to correct a specific problem within his technique, Narin wanted to construct a general measure of influence, applicable to all levels of aggregation and different types of entities. Apparently, different techniques and contradicting theoretical assumptions concerning the nature of the citation and the reference can go together smoothly.

To sum up, normalization, born out of the desire to obtain statistically more significant results, is at the same time a crucial way of connecting the citation with other signs or indicators. It facilitates the construction of indicators for science policy at every conceivable level of science.

5.5 Other signs of science: co-word analysis

In this chapter, the foundations and uses of citations and references in the specialty of scientometrics have been discussed. Scientometricians are like "heterogeneous engineers" (Callon 1986, Latour 1987), mixing various elements in their construction of representations of science. The incorporation of different representations of science into citation analysis, e.g. the validation procedures in cocitation analysis has been shown. This translation of various representations can, however, be carried even further by creating bibliometric representations of science nót based on the citation. This is what happens in co-word analysis, created by the French sociologist of science Michel Callon and his colleagues. According to this group, there is a fundamental problem with citation analysis:

> The study of citations limits the scope of the analysis to one of the numerous means used by an author to identify, mobilize or turn aside for his own profit earlier results, institutions and authors. (Callon, Courtial, Turner & Bauin 1983)

In this argument, the citing author is the central unit of analysis and the inversion implicated in the *SCI*, citation analysis and co-citation analysis is consequently rejected. Ironically, Callon et al. (1983) return to a representation of the literature discarded by citation indexing: subject indexing. Whereas Garfield rejected subject indexing as too rigid (and slow), preferring the newspeak of citation, Callon turns the flexibility of citation against its use: "Depending on the different disciplines, circumstances and audiences, he [the scientist] could have recourse to citation in different ways". Callon et al. (1983) prefer the old-fashioned subject indexing terms, precisely because of its rigidity:

> at least three factors tend to 'objectivate' the work of indexers: 1: the organizational stability of the documentation service; 2: the contact that the indexers have with the users of the documentation service when handling

> their requests for information; 3: the form of the scientific text itself with its double-mouth funnel-like structure which facilitates identification of the macro-terms.

This preference is partially inspired by the prevailing access to databases: the most important French scientific database *Pascale* did not include citations but did have a highly developed subject indexing system, whereas the inverse was true for the *SCI*[6].

This does not mean that the French would not have learned from the experiences with citation analysis. In more than one respect, co-word analysis is the analogue of co-citation clustering applied to a different database and with a different elementary concept. As before, "extension by analogy" is the applied procedure. What the citation is in citation analysis, is the indexing term in co-word analysis. Then, co-occurences are measured, just as in co-citation analysis: "two key words, i and j, co-occur if they are used together in the description of a single document" (Callon et al. 1991). Because of the never-ending possibilities of indexing terms, "It is clear that a simple counting of co-occurrences is not a good method for evaluating the links between co-words". Therefore, a normalized coefficient has been constructed which relates the co-occurrence of each pair of indexing terms to the general occurrence of each term, like the Jaccard coefficient in co-citation clustering. This is the *equivalence index*:

$$E_{ij} = \frac{C_{ij}^2}{C_i \times C_j} \tag{5.8}$$

where C_i is the number of occurrences of keyword i and C_{ij} is the number of co-occurrences of keyword i and j. This index is calculated for every possible word pair. On this basis clusters are constructed "with variable thresholds, characterized by the value of the first link refused, which is called the saturation threshold of the cluster. These thresholds are automatically determined in such a way that no cluster contains more than ten words". So, this clustering algorithm is based on a fixed upper size of the word clusters. Then these clusters are classified. If the division seems to be artificial, i.e. if one cluster is simply the continuation of another, a new round of classification is started.

> The first stage of the description of a network (we mean the whole network of words for a given file) is the identification of clusters, the description of the links that unite them, and the representation of their internal organization. We then have to characterize the morphology of the network as a whole, and the contribution of each of these clusters to its structure. (Callon et al. 1983)

After clusters have been generated in this way, indicators typifying them are constructed. Each cluster is defined by its centrality (the intensity of its links with other clusters) and by its density (the strength of the links that tie the words making up the cluster together). This double characterization of each cluster makes

[6]Michel Callon, Jean-Pierre Courtial, William Turner, Interviews, March 1995, Paris.

it subsequently possible to classify them into four quadrants and analyze the development of a network, or the difference between two different networks (which is an identical problem) in terms of the differences between these classifications. This means that clusters are analyzed on the similarities between the values of their densities and centralities. Clusters are moreover compared by computing the number of words they share, the so-called transformation index. Since this is the number of items in the intersection of the clusters, this is again a co-occurrence measure but at the higher, cluster, level.

As will be clear, co-word analysis, like citation analysis, rests upon several translations that are neither self-evident nor necessarily consistent. It takes the indexing terms to be adequate representations of the content of the scientific literature and is moreover based on the assumption that these terms remain stable over time and in different contexts. These co-occurrences are then translated into relative measures as input into clustering methods which have, as in co-citation clustering, several degrees of freedom. Different threshold criteria, for example, will lead to different clusters, as will other cluster criteria. Furthermore, the technique is based on relational analysis, like co-citation clustering, and does not take the relative positions of the words into account (for positional analysis of networks see Burt (1982)). This is borne out in the choice of relational indicators characterizing the clusters, only one of the many possible ways of discriminating between the clusters. Ultimately, co-word analysis resurrects the boundary between the method and reality (de-emphasizing the local context of the origin of the indexing terms):

> Co-word analysis, and in this it is in line with the sociology of traduction which gave rise to it, does not rely on a priori definitions of research themes. The subject areas identified are those which are constructed by different actors (researchers, engineers, ...) and which they define and transform in the course of their interactions. Co-word analysis considers the dynamic of interactions to develop as a result of actor strategies. (Callon et al. 1983)

5.6 A maze of indicators

> The utility of any particular indicator depends ultimately on the accuracy of the observations on which it is based, on the validity of the unstated assumptions by which it is accompanied, and on the logical consistency of the further processes by which it is reduced to operational form. (Ziman 1979, 261)

Thus, the citation representation of science is an intricate maze of indicators. It developed mainly through the extension of the sign citation, the juxtaposition and combination of the reference and the citation, and the combination with other representations of science. This development consisted of consecutive steps of translation, not necessarily consistent but still building on its own results. The citing and the cited dimensions of science have been thoroughly intermingled.

Translation also occurs whenever the citation representation is mixed with others. The citation culture has even permeated through a competing mode of bibliometric research like co-word analysis. Although based on the old-fashioned subject indexing, co-word analysis is in no way part of the pre-citation world. On the contrary, its methodology is deeply embedded in the established reality of citation analysis. Both the citation and the co-word representation of science are formal representations, contrary to the substantive science representations[7].

Apparently, the indicators used in citation analysis provide for a rather complicated representation of science. They do not reflect science straightforwardly. Far from simply measuring independent objects "out there", citation analysis selects and reconstructs objects on the basis of the citation index, on which it capitalizes. The heart of the matter is the semiosis of the citation, a symbolic inversion process. This transformation creates the basic properties of the citation network. The countability, universality, and self-referential nature of the citation are its central attributes. They make extensive new combinations possible as well as the simultaneous use of different, even supposedly incompatible, ways of measuring.

In this process of creating a new representation of science the realist and the constructivist perspective have frequently alternated. Scientometricians are able to use both discourses at the same time, as do scientists (Shapin & Schaffer 1985, Mulkay, Potter & Yearly 1983, Latour 1987) and sociologists of science (Ashmore 1989). In realist rhetoric, the scientometric indicators are depicted as being based on science itself. In constructivist utterances the creative act of making science indicators is stressed. During these activities, scientometricians are engaged in boundary work (Gieryn 1983), though not so much rhetorically as practically: deconstructing and reconstructing the boundaries between their constructs and the reality of science. Scientometricians thereby construct the reality of science *within* (or in terms of) the representation they have created. This is a general feature of scientifically constructed representations: not only do they represent a phenomenon by a combined reduction-reconstruction translation, but within this whole they also create boundaries between that which is labelled as deliberately constructed (like methods and artefacts) and that which is attributed purely to the phenomenon (results). This labelling can subsequently be swapped as we have seen in the evolution of the co-citation clustering methodology. Scientific representations generally indicate their own construction in a way that enhances their claims to be true. Scientometric representations are no exception.

Back to Contents Page

[7] This point will be discussed in more detail in chapter 8.

Back to Contents Page

Chapter 6

Rating science

6.1 Introduction

So far scientometrics has been analyzed as a so-called "data-driven" specialty. Without the *SCI*, scientometrics in its present form would not have existed. The sign *citation*, the foundation of a new system of signs of science, has shaped social opportunities for people to make a living in new ways (by producing citation indexes, performing citation analyses, and thinking up new indicators). In its turn, however, this has created a new demand for science and technology indicators. The field did not only develop on the basis of the "push mechanism" provided by the *SCI* and the citation sign system's unfolding features. A "pull mechanism" was at least as important, originating from the policy market for science and technology indicators. Statistical description of the research enterprise and evaluation of research and research programs have been the main uses of these indicators. The indicator market is itself a historical product, created by the breakdown in the early seventies of the virtually complete autonomy of post-World War II science on the one hand (Greenberg 1967, Blume 1974, Dickson 1984), and the new citation sign system on the other hand. It is highly doubtful whether scientometrics would have turned into a distinct social scientific specialty (instead of a hobby of some historians or merely one of many techniques of sociologists of science) without science policy as a market for its applications.

Funding bodies such as the US National Science Foundations (NSF) and the National Institutes of Health (NIH) have been the prime catalysts in inducing the creation of new performance indicators from the very beginning of present-day science policy. At an NIH meeting in 1955, Joshua Lederberg recalled having read Eugene Garfield's article in *Science* (chapter 2). The issue at hand was whether it would be possible to estimate the effectiveness of NIH's research funding. The same question was the starting point of the invention of co-word analysis. The French organization responsible for most French chemical research in 1976 wanted to know how good their researchers were and contracted Michel Callon to devise a proper method[1]. The idea of co-citation analysis was born in a comparable context at the American Physics Institute[2].

[1]Michel Callon, Interview, 20 March 1995, Ecole des Mines, Paris.

[2]Henry Small, Interview, 30 November 1993, ISI, Philadelphia

As the greatest power in science since World War II, the United States has been the main source of new developments in science policy in the Western world (Greenberg 1967, Blume 1974, Dickson 1984). The reports and actions of NSF and NIH profoundly influenced international science policy. In the field of technology assessment for example, the US Office of Technology Assessment OTA has been the model for many other national technology assessment offices. It is no coincidence that the *SCI* originated in the US (chapter 2). US science policy was also the context of the very first science indicators report (Board 1973) in 1972. Yet, historians of US science policy have paid scant attention to the development of these indicators (Greenberg 1967, Dickson 1984, England 1982, Golden 1988, HSTC 1980). This may itself be taken as an indicator of the lack of political clout of citation-based indicators: apparently they seemed irrelevant to scholars.

Postwar science policy in the United States was, until around 1970, based on the compromise reached in the debate about Vannevar Bush's report "Science: The Endless Frontier" (Bush 1945). The scientific communities were allowed to run their own affairs as they saw fit. The government was supposed to provide sufficient funds for science to grow. This would more or less automatically lead to applicable technology and hence prosperity and national security.

> The plain fact is that science has become the major Establishment in the American political system: the only set of institutions for which tax funds are appropriated almost on faith, and under concordats which protect the autonomy, if not the cloistered calm, of the laboratory (Price 1962)

Central co-ordination did not exist. In this respect, the National Science Foundation was given a limited task and would become in many respects less influential than the National Institutes of Health. These power relationships were forged in a five-year-long debate immediately after World War II between proponents of the autonomy of science and those adhering to the normal administrative standards, according to which "persons with responsibility for the disbursement of public funds should not be actively associated with the beneficiaries of those funds" (England 1982, 71,7). The proponents of a large degree of autonomy won, leading to the dominant philosphy of NSF according to which basic research was "self-coordinating" (England 1982, 338). The US military also financed a large body of undirected basic research. The Cold War Sputnik crisis in 1958, which initiated a redefiniton of much of American science policy and created a host of new institutions, amongst which was the House Committee on Science and Technology (HSTC 1980), did not basically upset the division of tasks between policy officals and scientists. Quality control stayed in the hands of the scientists themselves: after all who else could judge the often esoteric research results? As a consequence, assessing the state of affairs in a given scientific domain was based on the predominantly qualitative judgements of the experts in that domain. Quantitative indicators were mostly restricted to budget figures and personnel estimates.

During the sixties the need to justify budget requests for research projects more specifically started to emerge. Partly inspired by Derek Price's thesis of the exponential growth of science, and its imminent levelling off (Price 1961, Price

1963) the fear started to spread that federal support for science would stagnate. This was the reason the National Academy of Sciences created the Committee on Science and Public Policy COSPUP in which former presidential science advisor George Kistiakowski played the prime role:

> COSPUP was the creation of one of the grand and prescient statesmen of science, George B. Kistiakowski, who, while serving as presidential science advisor in the last two years of the Eisenhower administration, foresaw that the time would soon come when government would ask very hard questions about levels of support for basic scientific research. It cannot be said that science was granted a blank check prior to 1960. Far from it. But, though government did not grant the scientists all they sought, it nevertheless granted a great deal. The important point, however, is that what was given, was given on faith in the value of science, and not as a consequence of any systematic assessment of the place or value of science in national affairs. (Greenberg 1967, 159–160)

COSPUP main function was to "serve as a scholarly, dignified advocate for research by producing inventories of the scientific status of various fields and assessing the resources required to follow promising lines of inquiries" (Greenberg 1967, 160). This committee organized one of the first assessments of a scientific discipline, The Westheimer Survey of Chemistry (Westheimer 1965). From the beginning, it set out to demonstrate that "free, undirected basic research in chemistry was the source of much of the industry's prosperity" (Greenberg 1967, 161). In other words, the study was supposed to prove empirically what had been assumed in science policy since World War II. An important weapon in the study was reference analysis: the references to past research were measured to show the dependence of important technology on previously performed scientific research[3]. Whether the study was influential is questionable[4]. The Westheimer study is nevertheless an early example of the use of scientometrics by the relevant scientific experts. This type of citation analysis was by definition unable to upset the balance of power between the policy and the scientific community with regard to policy for science. The representation of science partly based by scientometric indicators was exclusively constructed by scientists.

During the sixties, this balance of power was gradually changed by the small science of science community which started to use the *SCI* as an important data source and also by the sociologists of science. They produced a number of case studies, often highlighting policy-relevant issues, such as the productivity of scientists (Britton 1964) or the structure of information exchange in psychology (Garvey & Griffith 1964). Sometimes these were very influential, such as the notion of the exponential growth of science. Indicators for policy however, were not routinely produced. These were still confined to the case studies they were

[3]A different study initiated in the same period by the Department of Defense, project Hindsight, concluded the reverse with respect to modern weapon technology and post-war physics research.

[4]In 1967, Westheimer said about his study's possible policy impact: "The effects are difficult to assess." (Greenberg 1967, 165)

part of. In 1972, indicators for science policy acquired a more independent status with the publication of the first *Science Indicators* report by NSF (Board 1973). The report devoted a small section to citation data prepared by Computer Horizons Inc., a company led by the physicist Francis Narin[5] (Narin 1976, 32). The NSB noted:

> There are certain relatively direct results of R&D which provide indicators for comparing the scientific and technical performance of nations. Primary among these are reports of research published in scientific and technical journals, citations of reports from these journals, and patents for new products and processes. (Board 1973, 5)

The main goal of this global citation analysis was the comparison of the scientific output of economically competing countries. The relationship between science and technology was the other main concern. The reports were an exercise in descriptive statistics of the science system at a high level of aggregation. The Science Indicators series inspired science policy officials abroad to create comparable documents. Whether they really influenced the direction of American science policy is, however, doubtful. Whether they affected quality control procedures in science is even more questionable. Evaluating scientific research stayed firmly in the hands of scientists themselves, embodied in the set of procedures collectively known as peer review (Chubin & Hackett 1990).

In countries other than the USA, *SCI* based quantitative science and technology indicators have played a more influential political role. They seem, moreover, to have been on the rise in recent years. In many European countries, and at the level of the European Union's scientific programs since the end of the 1980's, citation indicators have become more prominent. In Asia increasing attention to scientometrics seems to be developing. In India, for example, scientists are routinely citation-analyzed by the large bibliometric community of the country. In Australian science policy, the role of scientometric work seems to be increasing as well. This trend is not confined to government policy. The world's largest biomedical private funding trust, the Wellcome Trust, houses an active scientometric group which has become quite prestigious within the scientometric community. The world's largest scientific publisher, Elsevier Science, commissions citation analyses of its journals on a routine basis. One could possibly even speak of a "breakthrough" of these indicators in the last five years, and of a "coming out" of scientometrics as a regulatory science of science (Jasanoff 1990). As a consequence, scientists and scholars in the social sciences and the humanities seem to have become more aware of the existence and relevance of their citation scores.

The emergence of indicators in science policy in the Netherlands, the focus of this chapter[6], is illustrative of their paradoxical role in science policy in general.

[5] Narin had become involved with producing indicator reports for policy in the follow-up of project Hindsight, the Traces study (Narin 1969) in 1969. Francis Narin, Interview, 30 November 1993, CHI, New Jersey.

[6] I would like to emphasize that this chapter does not aim to treat the whole spectrum of the use of indicators in science policy. There already exists a vast literature on the way indicators should, and should not, be used. Neither do I aim to treat all policy relevant discussions and controversies

At first sight, quantitative indicators do not seem very important. Science policy decisions are generally based on political and technical arguments, not on quantitative evidence. Seen as instruments of power, the indicators certainly gave rise to a new group of experts. But this group has not wrested the decision-making procedures from the hands of representatives of the scientific community. Scientometricians' role is subordinate to the elites who have been in charge of science policy all along. Yet, the citation representation of science has steadily increased its presence in science policy documents in the Netherlands. The creation of a national observatory of science, "het Nederlands Observatorium van Wetenschap en Technologie" (van Raan, Soete, Beelen, de Bruin, Moed, Nederhof, Noyons & Negenborn 1994), grounded in the cooperation of the main Dutch scientometric group CWTS and the foremost econometric unit MERIT, is the apex of this development.

Gradually, the very definition of "science" and of "scientific quality" in science policy seems to have become affected by the citation representation of science, embodied in the scientometric indicators. This development derives its importance from the general evolution of Dutch science policy. It began after World War II with a "policy for science", in which the state funded and the scientists themselves decided what to spend the money on. Nowadays, power relations seem to have been turned upside down. A national foresight exercise in 1997 has led to the political formulation of scientific priorities to be pursued by universities and research institutes (Wouters 1996*b*). The scientometric indicators certainly did not cause this fundamental shift of their own accord. But, in a subtle way, they have contributed to the opening up of the bastion science once was.

This chapter probes into the introduction of indicators at the interface of science policy and the citation representation of science. The emergence of indicators in science policy in the Netherlands is exposed using available archives[7]. In this way, I hope to provide some insight into how the market for science indicators was created in the Netherlands.

6.2 Early Dutch science policy

The beginning of Dutch science policy can only be understood when examined in relation to the international discussion. The Dutch initially followed developments in other OECD countries (Blume 1986). The Dutch word for science policy "wetenschapsbeleid" was probably used for the first time in the draft law on higher education of June 1952 (OenW 1966, 1). From 1956 onwards a public discussion on the need of a national science policy arose, stimulated by the journal *Wetenschap en Samenleving* (Science and Society). The 1963 OECD report "Science and the policies of governments" strongly influenced the Dutch scene. On July 31, 1963, the development of an active national science policy was declared an offical

---

that have surrounded scientometrics. Worthwhile in its own right, this needs a study of its own and would greatly exceed the boundaries of this enquiry of the citation culture.

[7]Not all archives I wished to study were available, unfortunately. This is the reason some parts of this story are sketchy. See the appendix on archives and interviews.

goal of the new government (OenW 1966, 2)[8]. In the fall of that year a working group on science policy was created by the Minister of Education, Arts and Science Theo Bot[9], which was the basis for the creation of a new advisory body, the Advisory Council for Science Policy (RAWB)[10], following the British example of the Council for Scientific Policy (Blume 1986, 19). This council was to co-ordinate and stimulate the new science policy in which science, the social sciences and the humanities were to be treated on an equal footing[11].

The considerations leading to the RAWB bear the mark of the new science of science. Derek de Solla Price's thesis of the exponential growth of science (Price 1961) had a huge impact [12]:

> Many experts are of the opinion that science has grown exponentially for the last two to three ages, leading to a doubling in size in a few years. Today about 87 % of the scientists of all times would be alive. In about 30,000 journals some 600,000 articles are published every year, leading to a growth of 6 % of the 10 million treatises already produced[13]. (Diepenhorst 1966, 7)

At the first meeting of the interdepartmental committee of civil servants on science policy IOW[14] on 22 September 1966, Derek Price was extensively quoted by the chairman dr. A. J. Piekaar[15], concluding that "this exponential growth cannot go on forever"[16]. The threatening scarcity of funds for research was the main worry of early national science policy in the Netherlands.

---

[8] "Op het gebied van de wetenschappen zal de regering streven naar bevordering van een krachtig nationaal wetenschapsbeleid en van internationale wetenschappelijke samenwerking". (Cited in OenW (1966, 2))

[9] This was the "gespreksgroep voor de organisatie van het wetenschapsbeleid".

[10] "Raad van Advies voor het Wetenschapsbeleid". It had taken two years for the two Ministries of Education and Economic Affairs to agree on a policy with regard to applied scientific research (Kersten 1996).

[11] "Gespreksgroep en Parlement achtten het van groot belang dat geesteswetenschappen en maatschappijwetenschappen op gelijke voet als natuurwetenschappen in wetenschapsbeleid worden betrokken." (OenW 1966, 6)

[12] "Waarom een nationaal wetenschapsbeleid? Studies hebben uitgewezen dat het wetenschappelijk onderzoek en ontwikkelingswerk in de geïndustrialiseerde landen een exponentiële groeit vertoont." (OenW 1966, 2)

[13] "Vele deskundigen zijn van mening dat nu reeds twee of drie eeuwen de wetenschap een exponentiële groei vertoont, telkens in luttele jaren dubbel zo groot wordend. Vandaag zouden ongeveer 87 % van alle zich aan wetenschapsbeoefening door de loop der tijden gegeven hebbende geleerden aan het werk zijn. Er verschijnen in omstreeks 30 000 periodieken jaarlijks een 600 000 verhandelingen, die met 6 % de tot dusver reeds vervaardigde 10 miljoen verhandelingen verhogen."

[14] "Interdepartementaal Overleg voor het Wetenschapsbeleid".

[15] This meeting discussed the request for advice by the RAWB on the first national science policy budget, "het Wetenschapsbudget 1964–1966" which was to be published as part of the national budget for the year 1967 ("de ontwerpbegroting 1967").

[16] "Geleerden als De Solla Price hebben de groei van de wetenschapsbeoefening wetenschappelijk onderzocht en zijn daarbij tot merkwaardige uitkomsten gekomen. Niet alleen vertoont het aantal onderzoekers een regelmatige toename (een verdubbeling elke tien tot vijftien jaar), ook de kosten per onderzoeker nemen toe. De toename van het aantal eminente, bijzonder begaafde, onderzoekers verloopt echter langzamer dan die van het totaal aantal onderzoekers. Aangezien deze laatste weer minder is dan die van de totale kosten van het onderzoek zou hier de wet van de afnemende meeropbrengst zich doen gelden. Het is duidelijk dat een exponentiële groei, welke

6.3 Scientometrics within a funding body

Within funding agencies, the available budget became a practical concern. The principal funding body of physics research, the "Foundation for Fundamental Research of Matter"[17] FOM was the locus of the first scientometric activities in the Netherlands.

This organization was confronted with a decreasing budget growth from 1968 onwards. The main problem was the setting of priorities (le Pair 1969, le Pair 1974, 113). The foundation also feared that decreasing budgets would diminish its guiding role vis-à-vis research in the autonomous universities (le Pair 1969, 123). Physicist Cees le Pair started working at FOM in co-operation with the board to help solve the policy dilemmas in 1968:

> Just when I arrived the growth in FOM budgets started to decrease. I was trained as a physicist and actually never even heard of the term "science policy". But the board wanted me to work on it, because they were stuck in an impasse. Before then, the organization had grown with thirty to fifteen per cent per year in real terms. The main problem had always been how to find researchers good enough to spend the money responsibly. Now, all of a sudden, growth was gone and the organization had to reorient itself. I soon found out that those very smart physicists I was looking up to, didn't really know how to deal with distributing scarce money. Often they were simply talking nonsense. Science policy was not organized in a scientific way but was the victim of personal preferences and untested assumptions. And it was my job to create some order in this mess. So I started studying the structure and function of the research enterprise. I did an awful lot of reading at the time.[18]

The main question Le Pair tried to answer was: how do I distinguish the best researcher and the best proposal from others that are good as well? As it happened, the library of the State University of Utrecht had just acquired a set of the *SCI*. Le Pair thought he was the first to use them:

> They were not yet cut open, nobody had taken a look at them. They were still in the boxes in which they had been delivered by ISI. So, well, I decided to take a look at who was cited among these physicists I knew, and soon discovered that this varied highly. Not only from person to person, but also between subfields. Our plasma physicists, for example, were hardly ever cited though I knew they had a lot of prestige. But these guys were simply not interested in journal publications, their field moved so fast that they exchanged their results by mail with the few colleagues who could understand their results. So I soon learned the limitations of citation analysis as well.

In the United States the science of science had already built up a body of literature and expertise. The National Science Foundation had moreover started

die van het nationaal inkomen belangrijk overtreft, niet steeds kan doorgaan; ergens stuit ze op een plafond." (HOW 1966, 4)

[17] "Stichting voor Fundamenteel Onderzoek der Materie".

[18] Dr. Cees le Pair, Interview 16 April 1992, Utrecht, The Netherlands.

publishing the *Science Indicators* Reports in 1972. It was decided to send Le Pair abroad for a three month visit to NSF and other scientific institutions in 1970[19]. This laid the foundations for continuing international contacts between FOM and the bibliometric community in the United States. Although Le Pair shared a strong preference for quantitative data with his science of science colleagues in the States, he was more cautious than most about the use of citation data in science policy. From the beginning he stressed the need to balance the citation view with peer review, assuming they would prove to be complementary.

At FOM, a system of expert juries according to the DELPHI method had been set up to enable selective funding of physics research at Dutch universities and research institutes (le Pair 1976*b*). These juries were selected by the FOM staff, using a detailed database of physicists' specializations, publications, conference presentations and citation scores. "Finding the right experts is one of the most difficult problems in judging research quality." (le Pair 1976*b*, 18).

> The *SCI* was especially useful to find people who would otherwise be overlooked. It gave me ideas for potential jury members in specialties I did not know very well. On that basis I made lists for the board, who often reacted quite positive. Often they already knew the people involved, but simply hadn't thought of them beforehand. Only after they had endorsed my proposals several times, I told them I used citation data. Citation analysis was not very popular among physicists.

Nevertheless, Le Pair was given the freedom to organize scientometric studies, not as a main task but as a potentially useful sideline. It resulted in a small group of people producing a modest but steady stream of studies and publications on Dutch physics research.

> Returning from the States, I said we should start some more serious research into what you can and what you cannot do with citations. That resulted in two very fine studies, one on nuclear magnetic resonance research by Hans Chang (Chang 1975), presently director of FOM, and one on the electron microscope by Cees Bakker. Both aimed to find out who had been important in the scientific development. Citations turned out to be useless in the case of applied science like electron microscopy. Artefacts and patents are then predominant, not publications. This holds for all applied and technical science. We also developed a mathematical model with which you can really measure career mobility of scientists (Koeze 1974). I am still proud of those studies.

In the seventies, Le Pair's group was one of two scientometric groups in the Netherlands. The other group was led by the psychologist Marc de Mey at the

[19] This trip was paid for by the Dutch Ministry of Education (le Pair 1975, 179). The trip gave Le Pair the opportunity to compare US practice with his own use of citation data. At the request of NSF, he held a presentation about the potentials and pitfalls of citation analysis (le Pair 1970), comparing the results of the Cole brothers with newly undertaken Dutch scientometric studies on nuclear magnetic resonance research and the development of the electron microscope.

Catholic University of Tilburg and was less policy-oriented and more academically oriented. The Utrecht group developed international contacts and collaborated with the principal researchers in the science of science (Derek de Solla Price, Belver Griffith, Eugene Garfield, Michael Moravscik, Jan Vláchy and others). Le Pair was the only Dutch participant in the science of science conference "Quantitative methods in the history of science" which was held in Berkeley, California, from 25 to 27 August 1976 (le Pair 1976*a*). The group focused on physics in the Netherlands as its principal object of study (Chang & Dieks 1976). Their mathematical methods were borrowed from physics[20]. They considered citation indicators as subordinate to expert opinions about research quality and used them predominantly in combination with other quantitative data on scientific personnel, financial budgets and careers. FOM's executive board supervised the studies undertaken and even had the responsibility to approve them[21]. The board also approved publications[22].

The physicist-scientometricians at FOM began to distrust scientometrics as soon as it took on a life of its own:

> I always say that if you wish to use bibliometrics in science policy, you need guys who are in the midst of the field themselves. They must be people who will be visited by angry colleagues if they make mistakes. As soon as policy officials start working with citation data without knowing the field, disaster is certain.

6.4 Emerging Dutch science studies

The Utrecht scientometricians were mainly targeted towards physics research. Their international contacts were predominantly physics oriented. However they were also part of a national discourse, organized in the "Group of Science Researchers" GWO of the Inter-University Institute for Social Science Research[23]. Beginning in October 1975[24], this group met several times a year to discuss sci-

[20]For example, the decay rate of citations (Chang 1975, 143) is inspired by the mathematical description of radioactive decay. The group's "n-th partial mobility index" (Koeze 1974) seems to have walked straight out of a classical dynamics textbook.

[21]Chang (1975) was for example appreciated by FOM's executive board as "an important contribution to the field of the science of science" (Notulen U.B. 24-06-1975). After having participated in a UNESCO conference on "The evaluation in science and technology; theory and practice" in Dubrovnik from 30 June till 4 July 1980, Le Pair proposed to invite Derek de Solla Price. The executive board discussed this, wondering whether Price was the real expert, who should take the initiative to invite him and how this could initiate new developments in the Netherlands. In the end the board approved an invitation to De Solla Price for "one or more presentations on the science of science and a small symposion on three-dimensional memories" (Notulen UB - 19-08-1980).

[22]E.g. of an article written by Le Pair for the journal Universiteit en Hogeschool in 1980 (Notulen U.B. 02-12-1980).

[23]De groep wetenschapsonderzoekers GWO van de Stichting Interuniversitair Instituut voor Sociaal-Wetenschappelijk Onderzoek SISWO.

[24]The group was initiated by prof. dr. Egbert Boeker (a physicist), prof. dr. Gerard de Zeeuw and Peter Koefoed of SISWO; the latter became the group's secretary. From the very beginning the

ence studies presentations and projects. It constituted a loose form of communication aimed at satisfying the need for knowledge exchange within the specialty of science studies[25].

The group regularly discussed the application of quantitative methods. Citation analysis was discussed for the first time at a meeting on the 4th June 1976, when presentations were given by Dennis Dieks (FOM), Michael Moravcsik and T. Place (Catholic University of Tilburg) (Koefoed 1976). These were pioneering discussions: Moravcsik classified the different types of references found in scientific articles; Dieks argued for the use of a mathematical model based on a Poisson distribution of "decaying" citations to measure the impact of scientific articles within a discipline; and Place explained how, because of the Kuhnian revolution, he and his co-author Marc De Mey had arrived at scientometrics starting from the classical philosophy of science[26]. In November of the same year, a second discussion concerning the measurement of science was organized, at the occasion of the aforementioned Berkeley Conference on Quantitative Methods in the History of Science with Cees le Pair and the British sociologist Nigel Gilbert as the main speakers[27]. Gilbert presented a thorough review of the use of indicators to measure "the growth of science" (Gilbert 1976) and criticized citation analysis:

> Firstly, do citations represent a "roughly valid indicator of influence"? This question cannot yet be answered with any degree of confidence, for we do not know precisely why scientists typically cite others. Although some recent research has begun to shed light on the ways in which citations are used, and reasons for citations, this work has not yet reached a stage in which clear accounts of citation practice are available. (Gilbert 1976, 21)

He agreed with Cole (1970), though, that citations will serve "for some purposes as a roughly valid indicator of influence". Its uses to measure quality should be based on data on "confounding factors" which influenced the citation frequency apart from its intrinsic quality[28].

group was meant to improve the quality of science studies in a gradual and bottom-up approach. This choice was made because most researchers had already started their projects before the group started meeting and because of the heterogenity of the fields represented. (RAWB 1977, Bijlage 5).

[25] Apart from this group, four university institutes studied science. These were the Biohistorical Institute (Biohistorisch Instituut) in Utrecht, the Institute for the History of Mathematics and Inorganic Natural Sciences (Instituut voor geschiedenis der wiskunde en der anorganische natuurwetenschappen) in Utrecht, the Department of History and Social Aspects of the Natural Sciences (Vakgroep geschiedenis en maatschappelijke aspecten der natuurwetenschappen) at the Free University in Amsterdam and the working group Technology and Society in Leiden.

[26] "Bibliometrics as we use it has to be situated within the Kuhnian sociology of science. The bibliometric study of paradigms has two main themes: 1. scientific growth and 2. the structure of science." (Koefoed 1976, 5)

[27] The meeting was prepared with abstracts of several papers presented in Berkeley.

[28] At the meeting, Le Pair was a little more in favour of citation analysis than Gilbert. He pointed out that the exponential growth of science diminished the problem of mistaken definitions. He also brought forward that Gilbert's critical analysis applied to single papers: statistics of larger populations might meet several of Gilber's objections. Subsequently, Gilbert made clear that his paper was squarely in the anti-quantification camp for the Berkeley conference's sake. "My own position is more in the middle: it is very useful to measure things". (This account is based upon Arie Rip's notes made during the meeting (Arie Rip Personal Archive).)

The discussion of the potentials and pitfalls of citation analysis was part of the group's active involvement in the new sociology of science. Researchers presented draft chapters of books, prospective projects were discussed, and foreign visitors regularly attended the meetings. As was true of the new specialty as a whole, most meetings combined academic interests and policy orientations in the analysis of science as a social phenomenon. For example, on 23 March 1978 the meeting focused on Dutch science policy, a topic introduced by the officials responsible at the Ministry of Education and the Advisory Council for Science Policy[29] (Koefoed 1978*a*). The main question discussed was what could sensibly be studied by science researchers. The next meeting discussed a research proposal on Dutch science policy by two physicists from the Free University in Amsterdam (Koefoed 1978*a*, Koefoed 1978*c*). In May and September of that year, the meeting discussed research evaluation (Koefoed 1978*b*, Koefoed 1978*d*). These policy debates took turns with more fundamental theoretical discussions, for example with Ina Spiegel-Rösing on the draft introduction to a book which was by then already called "the SSTS bible" (Spiegel-Rösing & de Solla Price 1977).

It was a small group. In 1977, 34 scholars were members of the GWO, but only 14 were able to devote most of their time to science studies. A publication list made up in May 1977 by the RAWB staff counted 47 relevant publications by 24 different researchers (RAWB 1977, Bijlage 6). Of these, ten publications were scientometric in nature, nine of these written by the FOM scientometricians and one by Marc de Mey[30]. The majority of the group was mainly concerned with the impact of science on society and, from that perspective, with science policy.

The participants do not seem to have felt strongly that they belonged to one integrated field of research. STS was too fragmented, the scholars had varying research interests and a co-ordinated approach was lacking [31]. This was strengthened by the lack of institutionalization: there were no doctoral degrees to be had in science studies, and research programs in STS did not as yet exist at Dutch universities. This state of affairs was still felt to be unsatisfactory in the early eighties when the Advisory Council for Science Policy (RAWB) interfered with science studies in the Netherlands.

6.5 Science studies for policy

In Spring 1977, the Social Science Department of the Royal Dutch Academy of Sciences (KNAW) asked the RAWB if science studies should be stimulated in the Netherlands[32] (RAWB 1977, Bijlage 1). This was discussed in May and Oc-

[29] De Raad voor het Wetenschapsbeleid RAWB.

[30] The later scientometricians Loet Leydesdorff and Arie Rip were already on the list but they did not yet practice scientometric research.

[31] Personal notes Arie Rip; Koefoed (1978*c*).

[32] This request from the "Sociaal-Wetenschappelijke Raad van de Koninklijke Nederlandse Akademie van Wetenschappen" was the result of a discussion about science policy on 8 and 9 October 1976. Two considerations made science policy worthy of a study: science could help solve social problems and distributing scarce resources to science was becoming a bigger problem than it was before. It was concluded that a more scientific underpinning of science policy

tober, resulting in the Council's statement that "the science of science" was indeed important enough to begin developing the field (RAWB 1978*a*). The RAWB staff wrote a short report (RAWB 1978*b*) in February 1978 and discussed it with around 20 researchers. The RAWB interfered because this was not just a new field of study, but a specialty of pre-eminent importance to governmental science policy[33]. The RAWB staff was rather critical of the state of affairs in the Dutch science of science and deplored the lack of coherence of the field and of consensus among the researchers[34]. The council hoped that its action would end the impasse. In september 1978, the RAWB published its first advisory report on science studies (RAWB 1978*a*). It reformulated the goals of the specialty by putting "the study of the factors influencing scientific development", entitled "Science Dynamics", central. The report stressed in particular the need to reinforce "strategic research".

This report was discussed by the government in April of the next year, leading to the political decision to finance at least one full professor in the field of "science dynamics" from January 1980[35]. The RAWB was expected to take detailed measures in consultation with the researchers and universities. On 18 December 1979, the council appointed a committee[36] to prepare the creation of a university

---

was desirable. "Een nationaal wetenschapsbeleid, zoals dat zich ten zowel onzent als in andere landen het laatste decennium heeft ontwikkeld, kan in algemene zin op tweeërlei overweging zijn gefundeerd. In de eerste plaats kan het zijn grondslag vinden in de gedachte, dat bij een gegeven beperktheid van financiële middelen geen onbeperkte groei van de uitgaven ten behoeve van de wetenschap mogelijk is, en dat een wetenschapsbeleid er op gericht zal moeten zijn, de beschikbare fondsen zo goed mogelijk te besteden. In de tweede plaats kan een wetenschapsbeleid zijn fundering vinden in de gedachte, dat een gebundelde en gerichte aandacht voor het totaal aan wetenschappelijke inspanningen in een land ertoe kan leiden dat het niveau van de hieruit resulterende wetenschappelijke kennis en de gerichtheid op en de bruikbaarheid voor de oplossing van praktisch maatschappelijke problemen wordt verhoogd. In de regel zullen, naar mag worden aangenomen, beide factoren in een nationaal wetenschapsbeleid zijn aan te treffen. In beide gevallen zal daarbij steeds de gedachte op de voorgrond staan, dat wetenschapsbeleid (gunstige) invloed heeft op de aard en niveau van de op grond hiervan bereikte wetenschappelijke kennis. De vraag die dan rijst, is, welke invloedsrelaties er bestaan tussen wetenschapsbeleid en wetenschapsbeoefening en welke consequenties dit heeft voor de wetenschappelijke kennis en inzichten. Dit schept nu de grondslag voor de gedachte, dat wetenschapsbeleid zou verdienen zelf object van wetenschappelijke studie te zijn." (RAWB 1977, Bijlage 1)

[33]"dat de RAWB hierin het voortouw neemt vindt men alom gerechtvaardigd: het gaat hier immers niet om stimulering van een willekeurig wetenschapsgebied, maar om het vakgebied dat bij uitstek betekenis heeft voor de specifieke sector van overheidsbeleid waarop de adviserende arbeid van de RAWB is gericht" (RAWB 1978*a*, 1–2).

[34]"De grotere samenhang wil evenwel maar niet van de grond komen. Ondanks alle schone intentieverklaringen lijkt het in de praktijk onmogelijk om consensus te bereiken over wat men nu eigenlijk verder wil. Deels is dit te wijten aan de betrokken SISWO-coördinator die zich overwegend passief opstelt. Anderzijds staat er ook weinig druk op de deelnemers aan de groep om hun onderzoek een beetje af te stemmen op een gemeenschappelijke noemer, omdat hun onderzoek veelal gefinancierd wordt uit de eerste geldstroom." (RAWB 1978*a*, 3)

[35]"Minister Pais is daarom bereid ten minste één kroondocent zonder compensatie per 1 januari 1980 ter beschikking te stellen. De minister zal zich per brief tot de universiteiten en hogescholen wenden om na te gaan welke universiteit of hogeschool bereid is een vakgroep wetenschapsdynamica op te richten en daarvoor zelf een redelijke capaciteit ter beschikking te stellen opdat zodoende een groep onderzoekers van redelijke omvang — b.v. 5 à 6 manjaarequivalenten — binnen één universiteit werkzaam kan zijn."

[36]This committee consisted of: prof. dr. A. F.J. Köbben, dr A. J. Piekaar, prof. dr. H. A. Becker

department of Science Dynamics. This did not prove easy because of the lack of consensus among the researchers. Moreover, members of the council were not very satisfied with the level of Dutch science of science and seriously considered inviting a foreign researcher to organize the field (RAWB 1980*b*, 6)[37]. In the end the RAWB gave the green light, which lead to its second advisory report on science studies. Initially three, later four universities competed for the funding. In the end, a new team of science researchers at the University of Amsterdam won and founded the Department of Science Dynamics.

The result of this enterprise was the professionalization and institutionalization of science studies in the Netherlands as a policy relevant area, a regulatory science (Jasanoff 1990). It was different from the study of the history of science in that it was supposed to focus on post-World War II developments. It also differed from the "science and society" groups popular at that time in that it predominantly studied the potential of steering science, not its impact on society (RAWB 1980*a*, 5)[38]. Three topics in particular were outlined: the organization and institutionalization of research; scientific communication; and the influence from non-academic actors on scientific development (i.e. contract research and the effects of science policy itself). Although quantitative research was certainly included in the new science dynamics, it did not command any special attention. No specific mention of the merits of quantitative research was made in the documents or reports, nor were science and technology indicators mentioned in this context.

6.6 Indicators for policy

The development of science and technology indicators followed a different track, although it built partly upon research in the specialty of science studies. The OECD reports on Higher Education and Science Policy were especially influential. The first US *Science Indicators* Report (Board 1973) was also an important stimulus. As has been said, this initiative in the US was closely monitored by science policy officials in the Netherlands. And like in the States, the role of scientometric indicators was initially very restricted. Since 1974, Dutch science policy has been formulated in a separate policy document: the *Science Budget*[39]. Up to and including 1989, the only quantitative data used in this central document were so-called "input indicators" such as the amount of money invested in research and the number of scientific researchers at the universities (Wouters 1992*a*). This changed in 1990 and since that time bibliometric analysis has been part and parcel

and dr. K. H. Chang. (RAWB 1980*a*)

[37]Especially the chairman of the RAWB, Van Bueren, was highly critical: "Zelf is hij niet erg gelukkig met het advies. Het gehanteerde stimuleringsmodel zal naar zijn mening niet werken: wanneer momenteel geen bijzonder werk geleverd wordt zal dat volgend jaar ook niet het geval zijn." (RAWB 1980*b*, 6)

[38]"In het advies van de RAWB, dat uitgangspunt vormde voor de activiteiten van de Commissie Wetenschapsdynamica, wordt onder wetenschapsdynamica begrepen het onderzoek naar de ontwikkeling en stuurbaarheid van de wetenschap." (RAWB 1980*a*, 5)

[39]*Het Wetenschapsbudget*.

of the *Science Budget*. The Dutch government even formulated an explicit "indicator policy" in a 1991 policy document (W 1991). In 1994, the National Observatory of Science and Technology published its first *Science and Technology Indicators* report. The second one followed in 1996, and the series is supposed to continue with a frequency of a new indicators report every other year.

This political presence of the citation is the result of a long period of gestation in which the RAWB played the key role. From 1977 onwards, the RAWB started to devote itself to the promotion of quantitative science indicators for science policy. This was part of a general move towards more precise bookkeeping of the expenditures on scientific research and higher education. Its yearly reports show this development fairly well. In 1977, the RAWB pleaded for "quantifying the research effort"[40], in 1978 it developed an index to gauge the effective amount of research. One year later it published a study on scientific productivity. In 1982, the RAWB remarked that it would begin to develop science indicators, mainly to make comparisons with research investments and productivity in other countries. Over the next two years, the council published an influential report on the quality of health research (RAWB 1983*a*) as well as its first *Science and Technology Indicator Report* (van Heeringen, Mombers & van Venetië 1984). The RAWB's persistent campaign is the main cause of the present integration of scientometric indicators in Dutch science policy.

6.6.1 Research evaluation explored

The search for "output indicators" has been connected with the issue of evaluation from the very beginning. In 1978, the European Community organized a seminar on "The Evaluation of Research" in Copenhagen (Guzzetti 1995, 101). In the same year, a committee of the Academic Council [41] together with the National Science Foundation[42] organized a Dutch symposion on research evaluation[43]. It was meant to function as an introduction to this topic and was aimed at university managers and science policy officials[44]. Members of the aforementioned Group

[40]"Kwantificering van de onderzoeksinspanning."

[41]This was "de Commissie Algemene Vraagstukken Wetenschappelijk Onderzoek van de Academische Raad".

[42]This was "de Nederlandse organisatie voor Zuiver Wetenschappelijk Onderzoek ZWO".

[43]The symposion "Beoordeling van wetenschappelijk onderzoek; analyses van een beleidsinstrument" was held at May 26 1978 in Utrecht.

[44]"Met het symposium willen CAVWO en ZWO aan een ruime kring van wetenschapsbeoefenaars en anderen die in het kader van beleid in enigerlei vorm met onderzoekbeoordeling te maken hebben de gelegenheid bieden over het hierboven aangegeven thema van gedachten te wisselen. Daarbij wordt in het bijzonder gedacht aan bestuurders, beleidsmedewerkers, leden van wetenschapscommissies op de onderscheiden niveaus der universiteiten en hogescholen; personen uit de kring van de Academische Raad, zijn secties en commissies; personen uit de kring van ZWO, haar stichtingen en werkgemeenschappen; vertegenwoordigers van andere wetenschapsorganen; betrokken ambtenaren der verschillende ministeries, met name het ministerie van onderwijs en wetenschappen, respectievelijk dat voor wetenschapsbeleid; betrokken leden van de volskvertegenwoordiging; personen uit de bedrijfsresearch en research van andere particuliere instellingen; personen uit professionele verenigingen en wetenschappelijke bureaus." (de Ruiter & Vink 1978)

of Science Researchers GWO were involved in the preparations[45]. The first and main speaker, consultant and engineer John Boel, told the audience that research evaluation derived its relevance from a drive towards a more explicit and rational science policy. According to him, there was no way back to unfettered academic freedom, except perhaps for a small group of centres of excellence[46]. The symposion signalled a trend of "externalization"[47] and "formalization"[48]. In Boel's opinion, these trends represented challenges to the scientific enterprise, which could be partly met by grounding science policy itself in scientific research on research. The main point was, however, to keep research evaluation in the hands of the experts who knew the specific specialty.

This was also underlined by the Vice-Director of ZWO Professor Jolles. He firmly placed research evaluation within the domain of science proper with the thesis of "The Missing External Institution". Since research was routinely evaluated by peers, who were in their turn evaluated in the same way themselves, the group of evaluators and evaluated was according to him identical[49]. Jolles stressed that this should remain that way, since profound knowledge of the specialty involved was a sine qua non[50]. By making the distinction between "evaluation as practice" and "evalutation as object of research"[51], Jolles also formulated

[45]More specifically, these were the sociologist Henk Becker, the FOM official Cees le Pair, and the chemist cum sociologist Arie Rip. Also instrumental was Peter Tindemans, who would later become Director of Science Policy at the Ministry of Education.

[46] "De beoordelingsproblematiek ontleent zijn aktualiteit vooral aan het streven naar een (meer) expliciet en rationeel onderzoekbeleid. Daartoe is de ontwikkeling van allerlei beleidsinstrumenten (informatiesystemen, planningsysteem, beleidsruimte e.d.) vereist, maar uiteindelijk staat of valt het onderzoekbeleid met de mogelijkheid onderzoek expliciet en rationeel te beoordelen. Er is geen weg terug naar de "akademische vrijheid", behoudens wellicht voor een aantal zorgvuldig geselekteerde "centres of excellence" van beperkte omvang (en de toekenning van een beperkte "vrije ruimte" aan iedere onderzoeker)." (Boel 1978)

[47]"Meer en meer wordt het aksent verschoven van beoordeling van onderzoek door de wetenschappers zelf en naar wetenschappelijke maatstaven in de richting van beoordeling mede door niet-wetenschappers en naar andere dan wetenschappelijke maatstaven." (Boel 1978)

[48]"Meer en meer wordt gewerkt in de richting van formalisering van beoordelingsaspekten zoals kriteriastelsels, weegfaktoren tussen verschillende kriteria, voorgeschreven procedures etc. Deze vraag komt voort uit de vraag naar waarborgen tegen willekeur, naar mogelijkheden van beroep e.d." (Boel 1978)

[49]"Dat dit in de regel niet tot problemen leidt, kan wellicht in belangrijke mate worden toegeschreven aan de bij wetenschapsbeoefenaars aanwezige kritische gezindheid, die er o.m. toe kan leiden, dat men steeds ook twijfel aan de deugdelijkheid van eigen premissen en hypothesen en an de juistheid van eigen uitspraken blijft koesteren en dan ook op die grond steeds bereid zal zij begrip te hebben voor de twijfel en kritiek van anderen. In zoverre als kritiek op anderen en zelfkritiek gevoed worden door het besef dat de normen die hierin tot gelding komen gelijkelijk door allen gehanteerd mogen worden, blijven conflicten over het judgment by peers achterwege, ook daar waar de beoordelingen tot een verschillende uitkomst leiden." (Jolles 1978).

[50]"Indien de beoordelaar niet afkomstig is uit dezelfde tak van wetenschap als de beoordeelde, verzwaart dit de eisen die aan hem worden gesteld. In het algemeen kan dan worden gezegd, dat deze eisen kennis van het betrokken vakgebied en een breed en grondig inzicht in het wetenschapsproces omvatten. Aan dergelijke eisen voldoen in de regel slechts wetenschapsbeoefenaars of zij die tenminste een wetenschappelijke vorming hebben gehad." (Jolles 1978).

[51]"moeten twee zaken zorgvuldig worden onderscheiden: 1. beoordeling als bezigheid, als onderdeel van de taken van wetenschapsbeoefenaars; 2. beoordeling als object van wetenschappelijk onderzoek." (Jolles 1978)

a role for science studies[52]. Sociologists of science could help the assessing scientists in two ways: first they should check whether evaluation in the various specialties was performed in a comparable way, and second they could determine whether rating research performance was a profitable exercise[53]. The ZWO-director recommended in particular Robert Merton's sociology of science as basis for evalution[54]. This recommendation seems compatible with Jolles's emphasis on science's capacity to fulfill its own evaluation needs[55].

In other words, research evaluation and quality assessment should remain firmly in the hands of the scientific community. Hence, indicators were not assigned an independent function in research evaluation. This attitude, prevailing in the boards of ZWO and KNAW, was soon challenged by the central advisory council to the government in matters of science policy, the RAWB.

6.6.2 The RAWB medical research project

RAWB staff members were not by definition enthusiastic about citation analysis. On 2 May 1979, the GWO[56] discussed a project proposal by RAWB staff member Arie van Heeringen (Koefoed 1979) to measure the effect of mobility and age on scientific productivity[57] (van Heeringen 1979). Productivity was measured by weighted publication counts[58]. Citation frequency was not perceived as a very useful measure, since no appropriate citation index existed and because recent research could not be rated by citations (van Heeringen 1979, 9).

---

[52] "In het geval van beoordeling als object van wetenschappelijk onderzoek luidt het antwoord anders. Wij hebben hier dan te maken met een onderdeel van de "science of science". (Jolles 1978)

[53] "In de eerste plaats kan het van belang zijn te onderzoeken in hoeverre de beoordeling in c.q. door de verschillende wetenschappen op uniforme dan wel op verschillende wijze geschiedt, en, in het laatste geval, of en zo ja hoe deze verschillen samenhangen met de aard van deze wetenschappen c.q. met de aard van de door hen bestudeerde materie. In de tweede plaats kan het belangrijk zijn om na te gaan wat het rendement is van beoordelingsprocessen voor de aard en de kwaliteit van het onder invloed van die beoordeling staande onderzoek. (Jolles 1978)

[54] "In de eerste plaats — en dat is uiteraard een indicatie van onze eigen preferentie — zouden wij voor de studie van beoordelingsprocessen in het bijzonder Robert Merton's analyse-schema willen aanbevelen." (Jolles 1978)

[55] "Van beoordeling als bezigheid laat zich in het algemeen zeggen, dat zij met vrucht kan worden beoefend op een hoger generalisatieniveau dan het niveau waarop zich de beoordeelde projecten bevinden. Zij combineert dan een comparatief referentiepunt met een redelijke nabijheid tot de problemen. Deze gedachte doortrekkend kan men stellen, dat dan alleen de vraag naar de afweging van "wetenschap" tegenover bv. kunst of economische groei buiten de wetenschap aan de orde moet worden gesteld. Voor alle vragen van geringere reikwijdte ter zake van de wetenschap is zij zelf beter toegerust dan wie ook. (Jolles 1978)

[56] This group of science researchers was commonly involved in background discussions on the RAWB indicator projects. In December 1978, the GWO also discussed research evaluation on the basis of the CAVWO/ZWO symposion (Koefoed 1978*e*).

[57] The RAWB had formulated its approach on this topic in its 1976 yearly report. Van Heeringen's project was an attempt to further explore the factors influencing scientific productivity in order to promote it with specific policy actions.

[58] "Waar het voor dit project op aan komt is dat op grond van bovenbedoelde studies blijkt dat er, altijd voorkomende uitzonderingen daargelaten, in het algemeen een duidelijk verband is tussen aantallen publikaties van een onderzoeker en zijn (met andere methoden gemeten) kwaliteit." (van Heeringen 1979, 10)

In the next indicator project, however, citation analysis acquired a more central role. In November 1980, the RAWB hired medical scientist Henk Rigter to set up a project to assess the medical and health sciences[59] (Rigter 1980, Rigter 1981*a*), in anticipation of a governmental request. The project was made up of three parts: an input analysis, a global quality assessment, and an inventarisation of existing priorities in Dutch medical research including a review of possible new ways of priority setting in research funding. The main idea was to classify research according to disease, using the classification schedule of the World Health Organization. In a preliminary discussion within the Council, Rigter proposed the use of citation analysis:

> Giving a judgment on quality is a tricky affair. To be able to assess parts of such a broad field as medicine and health sciences, one needs a fast and practical method. Citation analysis seems most useful. Notwithstanding all objections against this method, its reliablity has been experimentally proven and repeatedly confirmed.[60]

As data sources both *Excerptica Medica* and the *SCI* would be used. A series of interviews with medical experts would enable validation and interpretation of the quantitative results. A number of RAWB members objected to the overambitious expectations of citation analysis. One member remarked that American researchers tended to overcite fellow countrymen, another stated that citation analysis was subject to so many problems that alternative methods of analysis should also be used. As a consequence, citation analysis was discussed more cautiously in the following project description:

> Citation analysis will be supplanted by other relevant analyses (patents; invitations; editorships of international journals). In the second phase of the project, many interviews will be held with representatives of the various specialties. This is the more important because citation analysis only indicates tendencies and must not be used in any absolute way[61] (Rigter 1981*b*, 3)

In August 1981, the Dutch government requested the RAWB to advise on "possible improvements in the manner in which priorities are set in health research in the Netherlands" (OenW 1981, van Onderwijs en Wetenschappen 1981,

[59]This was called "project Medisch-Wetenschappelijk Onderzoek".

[60]"Het geven van een kwaliteitsoordeel is een hachelijke zaak. Om een oordeel te kunnen geven over onderdelen van een zo veel omvattend veld als het MWO, is een snelle en praktische methode nodig. Van de beschikbare methoden lijkt de citeringsanalyse de best bruikbare. Ondanks alle bezwaren die men tegen deze (en alle andere) methode(n) zou kunnen opwerpen, kan gezegd worden dat de betrouwbaarheid van de methode van de citeringsanalyse experimenteel is getoetst en herhaaldelijk is bevestigd. (...) Per onderzoeksterrein (ziektebeeld) kan zo een maat verkregen worden van het belang dat door vakgenoten gehecht wordt aan Nederlands werk, in verhouding tot niet-Nederlands werk." (Rigter 1980, 2–3) (Rigter 1980, 2–3)

[61]"De citeringsanalyse zal aangevuld worden met eventuele andere relevante analyses (van patenten; invitaties; editorschap van internationale tijdschriften?). Het ligt uitdrukkelijk in de bedoeling tijdens de tweede fase van het project vele gesprekken aan te gaan met vertegenwoordigers van diverse MWO-onderzoeksvelden. Dit te meer omdat citeringsanalyses e.d. slechts tendenties aangeven en niet in absolute zin gebruikt mogen worden." (Rigter 1981*b*, 3)

Rigter 1986). The perceived need to limit the growth of the budget for medical research was one of the main motivations[62]. The letter signed by the Ministers of Science Policy, Health and Environment, and Education and the Sciences, explicitly recommended including citation analysis:

> Determining the possibilities to measure the actual output. One should first think of technologies like citation analysis which have already proved that they generate valuable insights in other domains. Other forms of output measurement should, however, also be studied, since the measurement of publication data can distort the picture of health sciences and medicine which are for a large part practically oriented.[63]. (van Onderwijs en Wetenschappen 1981, 2)

In his interim report in June 1982 (Rigter 1982), Henk Rigter explained to the Council that the input data were most difficult to gather. The budget management of the universities in general, and of the medical faculties in particular, did not allow for a detailed measurement of the investment in research. The output analysis, although very labour intensive, proved more promising[64]. The Council decided to use as many output indicators as possible (RAWB 1982*b*, 4). Citation analysis would figure prominently in the project. Scientific productivity was measured by counting the number of publications per department, excluding amongst others the publications in journals not part of the *SCI*[65].

The project's citation analysis measured the number of times a particular 1977 publication was cited in 1979 (Rigter 1981*a*, 4) and compared this with the citation frequencies of comparable articles from other West-European countries (Rigter 1981*a*, Bijlage 3)[66]. The extent to which Dutch medical scientists were members of

---

[62] "De prioriteitsstelling in het gezondheidsonderzoek in Nederland mag zich de afgelopen jaren in een toenemende belangstelling verheugen, zowel van de zijde van de onderzoekers en hun organisaties als van regering, parlement en de betrokken departementen, terwijl ook de belangstelling van de Nederlandse samenleving als geheel voor de gezondheidszorg en het daarvoor relevante onderzoek groeit. Hiervoor zijn verschillende redenen aan te geven. Allereerst betreft het hier een sector van onderzoek van vrij aanzienlijke omvang, hetgeen op zich al een nadere belangstelling rechtvaardigt. Maar ook bestaat het gevoel dat het mogelijk moet zijn de bestaande onderzoekcapaciteit in Nederland in de sector van de volksgezondheid op ruimere schaal dan tot dusverre reeds het geval is te betrekken bij het oplossen van velerlei vraagstukken in de volksgezondheid en de gezondheidszorg. Een belangrijke complicatie bij de beleidsvorming t.a.v. de volksgezondheid is dat het om macro-economische redenen niet langer mogelijk is de uitgaven in deze sector op dezelfde wijze te laten stijgen als in het verleden mogelijk was." (OenW 1981, 1)

[63] "Het nagaan van de mogelijkheden om de feitelijke "output" te bepalen. Hierbij valt allereerst te denken aan technieken als citatie-analyse e.d. die ook reeds op andere terreinen waardevolle inzichten hebben opgeleverd. Maar het is daarbij ook zinvol aandacht te besteden aan andere vormen van "output"-bepaling, aangezien de meetmethoden die zich op "papieren output" baseren een sterk vertekend beeld kunnen geven voor het gezondheidsonderzoek dat voor een belangrijk gedeelte ook praktisch gericht is." (van Onderwijs en Wetenschappen 1981, 2)

[64] "Deze in zekere zin unieke operatie lijkt echter wel vrij eenduidige conclusies op te leveren." (Rigter 1982, 2)

[65] "Uit de verslagen over 1980 hebben we uit deze lijsten geschrapt: uittreksels, ingezonden brieven, hoofdstukken in boeken, nederlandstalige artikelen, en voorts artikelen in tijdschriften met een dermate lage status dat zij niet in de Science Citation Index zijn opgenomen." (Rigter 1983*b*, 24)

[66] Rigter hired Peter Bakker to perform the citation analysis for a period of six months.

the editorial boards of medical journals was also studied, as one of the auxilliary methods to assess the quality of Dutch medical research (Bakker & Rigter n.d.). This method was also based on the *SCI*: only the 1168 most influential journals — measured by the ISI Impact Factor — from 48 fields of medical research were analyzed. Rigter and Bakker asked about 1000 distinguished non-Dutch medical scientists to mention outstanding Dutch medical scientists, "assuming that the number of times a person was mentioned would be a measure of his eminence as a scientist" (Bakker & Rigter n.d.). Again, the *SCI* was used: "many of these experts were chosen from the list of the 1,000 most cited contemporary scientists" (Bakker & Rigter n.d.).

The authors stressed the difficulties encountered in the measurement of scientific quality: "Quality is an elusive concept. No analysis can do more than give an impression of the influence and appreciation of Dutch researchers"[67]. Yet, they had an overall positive assessment of the possibilities of citation analysis: "The number of citations can be seen as a measure of the influence of a publication in the scientific community"[68].

Citation analysis was applied within the context of the classification of medical research according to pathologies. The Dutch publications about specific diseases were compared with a representative sample of comparable foreign publications. The focus was not on the individual researcher but on groups and specialties. The foreign sample represented "the world's average" within the disease-related specialty. The project also used a combined measurement of citation frequency and journal impact factors. The impact factor was considered as the world average and subtracted from the publication's citation frequency. In this way, the deviation from the average was obtained. Most differences found were, however, not statistically significant because of the huge variation between articles in disease-related specialties[69]. The editorial boards of the medical journals were also analyzed, amongst other things to find out if "editorial position" might be a useful science indicator.

The result was a ranking of health research departments based on a composite of the indicators used. The most reliable measures were, according to the draft conclusions presented to the RAWB in March 1983 (Rigter 1983*b*, 2), the citation

[67]Daarbij beseffen wij dat kwaliteit een moeilijk grijpbaar begrip is. Geen enkele analyse zal veel verder kunnen komen dan het geven van een indruk van de invloed van, en de collegiale waardering voor persoon en werk van Nederlandse onderzoekers, en dat is wat deze studie beoogt" (Rigter 1983*b*, 3).

[68]"Het aantal malen dat een publikatie geciteerd wordt in de internationale literatuur, kan men zien als een maat voor de invloed die de publikatie heeft op de wetenschappelijke gemeenschap. Invloed veronderstelt kwaliteit, hoewel beide begrippen niet hoeven samen te vallen. Er zijn talloze onderzoekingen die het nut van citatie-analyses bevestigen. Dat neemt niet weg dat er ook een aantal mogelijke bezwaren aan dit soort analyses kleven. De eventuele nadelen zijn breed uitgemeten in het rapport van de Verkenningscommissie Biochemie, waarnaar wij kortheidshalve verwijzen. Met de Verkenningscommissie zijn wij van mening, dat, bij een juist gebruik van het middel van de citatie-analyse, de voordelen opwegen tegen de nadelen. Bovendien hebben wij de citatie-analyse op een bijzondere wijze toegepast, om de verstorende rol van de nadelen te beperken." (Rigter 1983*b*, 30)

[69]"Dit 'hoger' of 'lager' geeft tendensen aan, die echter zelden statistisch significant zijn, door de grote variatie tussen artikelen onderling." (Rigter 1983*b*, 34)

frequency, the number of times a group was mentioned by foreign experts, and the judgment of Dutch experts. The number of editorial positions was perceived as a weak measure. The measures correlated strongly[70].

In February 1983, Henk Rigter sent a draft of two reports to the Council, to be published in RAWB's series of "Background Studies", the first (Rigter 1983*a*) dealing with investment in medical and health science, the second (Rigter 1983*c*) presenting the analysis of the quality of Dutch research in these areas. In an earlier meeting the RAWB had decided to publish these two studies as soon as possible because of the debate on budget cuts and reorganizations at the universities[71]. These two studies were, however, only meant to be the first half of the operation. A third background study would discuss methods of prioritizing medical research. The scientific criteria would be derived from the output analysis. How the societal criteria could be developed was less clear[72]. A fourth study was supposed to wrap up the arguments in the preceding three studies and would be the foundation for the Council's prospective advisory report on priorities in medical and health sciences. The two additional studies would underpin a more general operation to develop priorities and posteriorities in Dutch science policy. The government asked the RAWB to develop a general method of determining these priorities (RAWB 1983*g*, 3). The medical project would be the first exercise in this difficult area.

As it turned out, the second half of this operation did not make it beyond the stage of draft texts because the RAWB was not satisfied with the results. The pressure exerted by the heated discussion in the months before and after publication of the RAWB advisory report to the Dutch government probably contributed to this outcome. In 1982, Henk Rigter wrote an 87 page first draft background study (Rigter 1983*d*), in which he discussed a host of possible criteria ranging from research costs, the steerability of scientific research, disease-related suffering, mortality rates, and scientific quality. In general, he distinguished criteria based on health policy and on science policy, together forming a criteria quadrant. It proved difficult, however, to agree upon a methodology. The need to publish the RAWB advisory report to the government became more urgent than the publication of a neat background study. Council member Hans Galjaard, a prestigious medical researcher in the Netherlands, wrote a memorandum about ways to prioritize medical research (RAWB 1982*b*), which partly deviated from Rigter's approach, to speed up publication of the RAWB report. Based on the discussion of this memo, the staff wrote a draft report proposing a three-way procedure: first determine priorities in health policy; second determine how research can contribute to these priorities; third determine priorities ánd posteriorities (RAWB 1983*c*, 16–19). The fourth step would be implementation and its translation into research programs and strategies by researchers and scientific in-

[70]"De citatie-maat en het oordeel van buitenlandse en Nederlandse deskundigen sporen doorgaans goed met elkaar. Het komt zelden voor dat het werk van een groep goed geciteerd wordt, maar niet wordt geprezen door de buitenlandse of Nederlandse deskundigen." (Rigter 1983*b*, 2)

[71]This was the so-called "taakverdelings- en concentratie-operatie" (Hutter & Rigter 1983, 2).

[72]"De wetenschappelijke criteria kunnen worden ontleend aan de output-analyse; voor de maatschappelijke criteria evenwel moeten nog methoden worden ontwikkeld" (Hutter & Rigter 1983, 2).

stitutions. In June 1983 the Council members were still not satisfied with the draft texts (RAWB 1983*g*)[73]. A major rewrite (RAWB 1983*b*), with the help of Council member Hans Galjaard (RAWB 1983*e*, 2), was still found wanting. A main point of contention was the lack of clear criteria that could be applied consistently, as well as the limitations of citation analysis. In this meeting, the heat of the public debate made itself felt.

The output analysis was published as an issue of the RAWB's series of background studies[74] (Rigter 1983*c*) in August 1983. A few months earlier, however, it was leaked to the press, after the RAWB had sent a confidential preliminary draft to medical faculties and government departments[75]. The press concentrated on the report's ranking in which the medical faculty in Rotterdam was top (ANP 1983, Anonymous 1983*b*).

The quality assessment performed by the RAWB was the first such exercise in the Netherlands which threatened to have real-world consequences. The medical faculties had to cut their total budget by 88.9 million guilders within the next four years (Rigter 1983*d*, 85). They were therefore forced to discuss a more efficient division of tasks. The research groups which had ended in the lower region of the RAWB ranking had every reason to fear the consequences, notwithstanding the statements in the RAWB report that a low citation score did not necessarily mean low quality of research. For example, in June 1983 the RAWB drew the conclusion that the medical faculty at the Free University in Amsterdam could in theory be abolished, since the overall quality of its research was clearly lagging behind other faculties and another high quality university teaching hospital already existed in the Amsterdam region (Anonymous 1983*a*). It did not mean that the RAWB was uncritical of government policy. On the contrary, the council partly agreed with the widespread criticism in the academic community that the government threatened research quality with its strong budget cuts. But the RAWB stressed, more than most university representatives, that these cuts made the protection of high quality research even more imperative and hence also quality assessments such as its medical project. RAWB chairman Professor Henk van Bueren stated that the Council found it imperative to protect high quality research in the era of budget cuts[76].

With its report (Rigter 1983*c*), the RAWB certainly created a stir. This was visibly present in the report of the Council meeting mentioned above on the first of July, 1983 (RAWB 1983*e*). The meeting discussed the first draft of its advisory

[73] At this time, the RAWB still planned to publish its third background study on methods of prioritizing.

[74] "Serie Achtergrondstudies RAWB".

[75] The University of Amsterdam weekly *Folia Civitatis* got the scoop. It reported that a confidential working document of the RAWB had been sent to medical faculties. The document would be used in discussions about budget cuts ("taakverdelingsoperatie"). The national press followed it up.

[76] "dat de Raad het in een tijdperk van bezuinigingen van groot belang acht dat het werkelijk goede onderzoek wordt gehandhaafd en beschermd. Daartoe is evenwel tijdige identificatie van dat goede onderzoek nodig. (...) Een wetenschappelijk prestatieoordeel mag en kan aan dit totaal in de huidige tijd niet langer ontbreken, en is ook niet meer bedreigend dan welke andere beoordelingscategorie ook. Het is alleen nieuw, en veroorzaakt daarom wellicht commotie." (van Bueren 1983*b*)

report on priorities in health research, which would be based on Henk Rigter's study and was to be published in October that year (RAWB 1983*a*). A few council members stressed the need for especially careful editing of the final text, given the sensitivity of the subject. The meeting even decided to contact the State Prosecutor[77] and to ask advice from a public relations expert on the presentation of the conclusions of the report (RAWB 1983*e*, 2,5). The members of the council had differing opinions on the merits of the citation analysis used. Two of them stressed that social science should not be measured in such a way. The meeting concluded that different sciences should be evaluated in different ways[78]. Most members also disagreed with a perceived tendency in the draft report to make detailed statements about specialties; they preferred a more global judgement. It was decided to create a more subtle set of criteria by clustering priorities and posteriorities into eight groups (instead of the four originally proposed) and to stress its comparative (in constrast to absolute) nature.

In August 1983, the RAWB published its revised report on priorities in health research (RAWB 1983*a*). It stressed that quality assessment could not be absolute:

> The statements by the Council should not be read as absolute conclusions; they have to be interpreted against the background of the analyses performed which only lead to a rank order[79]. (RAWB 1983*a*, 6)

This rank was the basis on which to identify potential growth areas as well as scientific specialties in which the Dutch government was advised to spend less. Three criteria were crucial:

a The necessity of research based on priorities in health policy[80];

1. The necessity of research on the basis of impressions about research quality[81];
2. The amount of existing research[82]. (RAWB 1983*a*, 21)

This led to a matrix of eight cells, ranging from research that needed to be stimulated up to and including research that should be ended. Many of the recommendations were not directed at the central government but at the relevant university or research authorities: a centralized direction of medical research was

[77]"De discussie naar aanleiding van dit punt leidt tot de afspraak de MO te verzoeken om vrijwaring c.q. verdediging door de Landsadvocaat, en het advies niet uit te brengen alvorens daaromtrent garanties zijn verkregen." (RAWB 1983*e*, 2)

[78]"Gezien de discussie zal in het advies duidelijk worden geëxpliciteerd dat verschillende wetenschappen verschillend moeten worden beoordeeld, en zal waar nodig worden gewezen op de beperkte reikwijdte van de gehanteerde criteria." (RAWB 1983*e*, 3).

[79]"Niettemin kan aan uitspraken van de Raad geen absolute waarde worden toegekend; zij dienen steeds bezien te worden in het licht van de gebruikte analyses, die slechts leiden tot een *rangorde*. (RAWB 1983*a*, 6)

[80]"de wenselijkheid van het onderzoek, op grond van mogelijke prioriteiten in het volksgezondheidsbeleid"

[81]"de wenselijkheid van het onderzoek, op grond van indrukken over de kwaliteit van het bestaande onderzoek (als verwoord in Achtergrondstudie nr. 2)"

[82]"de omvang van het bestaande onderzoek (beschreven in Achtergrondstudie nr. 1)"

not advocated by the RAWB (RAWB 1983*a*, 26). The report did not hesitate to recommend restructuring of institutions, for example the Royal Institute for the Tropical Zones[83] (RAWB 1983*a*, 32), or decreasing investment in specialties of dental research (RAWB 1983*a*, 54) and skin diseases (RAWB 1983*a*, 59). Social medical science was not valued highly[84]. This had been a topic of debate within the RAWB earlier, and the advisory report stressed that the methods used might not be applicable to all fields of research[85].

Notwithstanding these nuances, the RAWB had produced a remarkable science policy document. It was strongly based on data derived from the *SCI* and, moreover, on a new combination of citation analysis and expert opinion. The latter was no longer the exclusive source of quality assessment of research. The report promoted itself as an exemplary study in the development of a methodology to determine priorities in scientific research. And lastly, it applied this methodology to propose very material priorities and posteriorities in health research, although the RAWB itself was not an expert panel in the specialties it assessed.

This was unheard of in Dutch science policy. Not coincidentally, the RAWB organized a special meeting with representatives of the central university boards[86] as the first occasion to present the report. The main purpose was to discuss the methods used to prioritize research domains and groups. This had been made more urgent by the invitation by the Minister of Education and Sciences to the RAWB, in February 1983, to engage in the more general exercise of determining priorities and posteriorities in Dutch scientific research as a whole (RAWB 1983*d*). The RAWB president stressed the special position of the RAWB as an independent body of advice to the government, not guided by its own political or scientific interests[87]. He moreover underlined the careful use that was made of scientometric indicators, which according to him resulted, although not perfect, in reliable conclusions[88].

Nevertheless, the RAWB report as well as the RAWB medical study were strongly criticized by many amongst whom were the former secretary of the med-

[83] Het Koninklijk Instituut voor de Tropen.

[84] "Het sociaal-geneeskundig, psychologisch en sociologisch onderzoek met betrekking tot de geestelijke gezondheidszorg is in Nederland in kwalitatief opzicht in het algemeen nog onvoldoende tot ontwikkeling gekomen."

[85] "Voor een goed begrip van bovenstaande opmerkingen over het sociaal-wetenschappelijk onderzoek binnen het onderhavige gebied, herhaalt de Raad dat de methodiek die gebruikt is voor de beoordeling van de kwaliteit van het Nederlandse gezondheidsonderzoek niet voor alle velden even geschikt hoeft te zijn."

[86] The "Colleges van Bestuur".

[87] "De Raad heeft getracht zo zorgvuldig mogelijk te werk te gaan, maar realiseert zich dat het kiezen en wegen in de practijk onvermijdelijk mede berust op subjectieve beoordeling. Dit geldt zowel voor de kwaliteit van het onderzoek, als voor de keuzen te maken voor de gezondheidszorg. Juist vanwege deze onvermijdelijke subjectiviteit acht de Raad zichzelf geschikt voor deze taak, omdat hij een ongebonden orgaan is en niet geleid wordt door politiek of wetenschappelijk eigenbelang." (RAWB 1983*d*, 5)

[88] "De kwaliteit van het bestaand onderzoek is uitvoerig onderzocht in achtergrondstudie 2, waarvan U allen reeds eerder concepten heeft gezien. De gebruikte criteria publicaties, citaten (per groep), redacteurschappen, peer oordelen van vooraanstaande onderzoekers leveren goed met elkaar kloppende gegevens; daardoor ziet de Raad het resultaat als zo betrouwbaar mogelijk — perfect is het niet en kan het ook niet zijn." (RAWB 1983*d*, 7)

ical council of the Academy of Sciences, Professor Van Bekkum[89]. Van Bekkum had chaired a quality assessment of medical research in 1973 on behalf of the Academy of Sciences. This report had included a quality assessment by medical experts. He had fully informed the RAWB staff of this earlier exercise. The main difference between the two reports was the position taken up by the experts in those fields evaluated. Whereas Van Bekkum's report had been based on the experts involved, he said the RAWB had only consulted experts superficially[90]. As Van Bekkum did not agree with the method used by the RAWB, he argued against their conclusions, accusing the RAWB of causing harm to medical specialties[91].

The criticism was shared by a number of low-rated departments. For example, the Department of Haematology at the University Hospital in Nijmegen, in response to the confidential first draft of the RAWB report, sent additional information explaining why its publications did not rank high in the *SCI* as well as why the staff did not have the time to take up many editorial positions, naming the foreign experts they were working with (Haanen 1983). The important national medical journal "Medisch Contact" published a critical comment written by its editor-in-chief Van Es, calling the RAWB report a budget cut report[92] and stating that clinical research had been grossly undervalued (van Es 1983). The only medical researcher who was member of the RAWB, Professor Hans Galjaard, went into the offensive at the 1983 congress of the Royal Dutch Medical Association[93], castigating his fellow physicians. According to Galjaard they were afraid of setting priorities[94].

---

[89]Van Bekkum was secretary of the "Raad voor Medisch-Wetenschappelijk Onderzoek (RMWO) of the Royal Dutch Academy of Sciences (KNAW) from 1970 until 1980.

[90]"Het belangrijkste verschil tussen de evaluatie van de RAWB en die van ons in 1973 is, dat de RMWO uit acht personen bestond die geselecteerd waren uit het medisch wetenschappelijk onderzoeksveld. Over de vakgebieden, die onze raadsleden niet zelf konden beoordelen (en voorzover zij vonden, dat zij hun eigen onderdeel niet goed genoeg konden beoordelen), werd het advies gevraagd van een of meer "meest in aanmerking komende deskundigen voor het betreffende deelgebied" aan de hand van een uitvoerige instructie. Die rapporteurs werden uitgenodigd hun vakgenoten bij de medische faculteiten te consulteren, al weer aan de hand van een aantal door ons uitgewerkte vragen (waarover het aantal publicaties in internationale tijdschriften). De RAWB heeft slechts één lid uit het medisch wetenschappelijke veld en heeft bij het consulteren van deskundigen buiten de Raad een oppervlakkige beoordeling in plaats van een onderbouwd advies ingewonnen." (van Bekkum 1983)

[91]"Nu ministers begrijpelijkerwijze alle argumenten aangrijpen om minder geld uit te doen geven, kan een negatieve beoordeling voor een vakgebied zeer ernstige gevolgen hebben. Wanneer het dan ook nog onzorgvuldig onderbouwd blijkt te zijn, bijvoorbeeld door het eindoordeel *niet* te toetsen aan een aantal toonaangevenden uit dat vakgebied, veroorzaakt dit terecht commotie en zelfs verontwaardiging." (van Bekkum 1983)

[92]"Snoeirapport."

[93]Koninklijk Nederlands Medisch Genootschap KNMG.

[94]"Helaas heeft die politiek nog niet de weg gevonden naar prioriteitenstelling en nu komt het interessante van onze volksaard: er zijn drie bewindslieden die vragen aan de Raad voor Advies voor het Wetenschapsbeleid instrumenten te geven voor die prioriteitsstelling, en die raad komt met methoden daartoe. De achtergrondstudies van dat rapport zijn door eenieder geaccepteerd en oogstten heel weinig kritiek toen ze drie maanden geleden voor de buitverdelingsoperatie circuleerden. Nu zijn ze opnieuw naar buiten gebracht en nu moppert iedereen er over. Want de Nederlandse wetenschapsman heeft nog niet geleerd bescheiden te zijn en te kijken naar iets dat als oordeel over hem wordt uitgesproken om vervolgens te zeggen: 'Goh, ik doe het kennelijk

All in all, the RAWB received around twenty letters critizing its report, amongst others one signed by 77 cancer researchers and the organized dental researchers. As RAWB President Henk van Bueren wrote in Medisch Contact:

> Seldom has an RAWB report created so much excitement[95].

Van Bueren pointed out that the background studies had provoked little reaction after their publication in March 1983. Only after the RAWB report had been made public did people become angry with the methods used to rate scientific quality. Van Bueren's publication (van Bueren 1983*a*) was part of an offensive mounted by the RAWB, following up on Galjaard's speech at the national physician's congress. On 2 December 1983, the RAWB discussed the letters they had received and how to respond to the sharp criticism (RAWB 1983*f*). The main problem was how to respond in such a way that the Council would not be seen to apologize or be on the defensive. One member proposed convincing science editors to pay special attention to citation analysis in the newspapers[96]. A good deal of the criticism was aimed at the method of citation analysis. The secretary of the RAWB, Wim Hutter[97], summarized the reactions to the council (Hutter 1983*a*). He distinguished the following points of contention:

- "citation and publication analysis is always retrospective and cannot map new developments". "True", wrote Hutter, "this is the reason we also asked expert opinions of some hundred distinguished researchers."
- "citation and publication analysis can only be applied to fundamental research and not to applied research". "This is the reason we used several differing criteria", stipulated Hutter, "and explicitly included applied research in our talks with experts, among whom clinicians."
- "not all journals have been included". "For reasons of quality assurance we have only used *SCI* journals", replied Hutter.
- "by choosing the journals some fields of medical research have been disadvantaged". Hutter: "patently untrue. We have used several analyses and

niet zo goed, niet als persoon, maar met mijn groep en in mijn discipline; hoe kan het beter?' Nee, dat beetje energie dat wij nog over hebben verprutsen we door met elkaar te zeuren over details omdat we bang zijn dat de echte waarheid naar boven komt." (Galjaard 1983, 1378)

[95]"Zelden heeft een advies van de Raad van advies voor het wetenschapsbeleid (RAWB) zoveel opwinding veroorzaakt als het onlangs aan de regering uitgebrachte rapport "Prioriteiten in het gezondheidsonderzoek". Ook in uw blad zijn enkele reacties gepubliceerd, in één waarvan zelfs termen als 'onkunde', 'misleiding' en 'tendentieuze aanbevelingen' worden gebruikt. Dit versterkt overigens onze indruk dat er in veel gevallen eerder sprake is van emotionele uitingen dan van serieuze kritiek op de door de Raad gehanteerde uitgangspunten, analyses en beleidsaanbevelingen." (van Bueren 1983*a*)

[96]"Hij stelt voor om enkele deskundige wetenschapsredacteuren over te halen een themakatern of iets dergelijks te maken over "citation analysis" en daarbij het rapport van de Raad als voorbeeld te gebruiken. Daarnaast zou de voorzitter in een medisch vaktijdschrift een uitgebreide "letter to the editor" kunnen plaatsen." (RAWB 1983*f*, 2)

[97]Henk Rigter had already become secretary of the Health Council and was no staff member of the RAWB any more.

the main one is based on *all* articles published by Dutch researchers in 1977. The critics are misreading the report and focus only on one of the additional analyses."[98]

- "the citation frequencies have been miscalculated". "This has happened with one research group only", wrote Hutter, adding that this group measured its citation frequency in the wrong way itself.
- "the choice of journals is wrong". Hutter: "Excerpta Medica is the most reliable database in existence and our choice has not been criticized by any of our discussion partners, among whom several of our critics."
- "the content of the expert interviews kept confidential". Hutter: "Yes and this will remain the case."[99]

An important principal point was also raised. The very competence of the RAWB was disputed because it only had one or two medical researchers in its midst[100]. Wim Hutter emphasized that experts always also have special interests and that the task of the Council was to weigh the various interests against each other. Hence it was not a weak but a strong point that the Council had few experts in its midst[101].

This is a crucial point with regard to scientometric and bibliometric indicators. The RAWB wished to promote the use of these indicators, in combination with expert opinions, to redress the balance of power over scientific research. Research should be evaluated by institutions with no vested interests in the field concerned. At the same time, the RAWB did not advocate the exclusive use of indicators at the expense of expert opinion. A tophit of the 95 most cited medical professors, published in the daily newspaper NRC/Handelsblad (van

[98]Hutter was adamant about this: "Het is duidelijk dat iedere kritiek op onvolledigheid e.d. die niet de integrale viervoudige opzet van de analyse tot uitgangspunt neemt, geen hout snijdt. Aangezien alle kritiek totnutoe slechts één van de additionele analyses tot uitgangspunt neemt en de overige, inclusief de hoofdanalyse buiten beschouwing laat, snijdt geen van de totnutoe ingebrachte kritiek hout." (Hutter 1983*a*, 4)

[99]"Zeker, en dit zal zo blijven. De ervaring is dat onderzoekers slechts dan openhartig over elkaars werk oordelen als ze weten dat dit niet aan de grote klok wordt gehangen." (Hutter 1983*a*, 5)

[100]The RAWB had one active medical researcher as a member, Hans Galjaard, and one former brain surgeon who was topleader of Elsevier, Pierre Vinken.

[101]"Met dit verwijt heeft de Raad vaak te maken, omdat hij niet is opgezet om een orgaan van deskundigen te zijn. Deskundigen vertegenwoordigen echter ook altijd belangen en het is juist de taak van de RAWB deze belangen tegen elkaar af te wegen. Hij doet dit door zich bijvoorbeeld door deskundigen te laten voorlichten of deskundig materiaal bijeen te brengen (achtergrondstudies). Ervaringen met deskundigencolleges (verkenningscommissies, het medisch cluster van de TVC-operatie) laten zien dat deze vaak niet tot echte prioriteringen (kunnen) komen, laat staan posterioriteiten aanwijzen. Voor het aangeven van rangordes, waarbij ook de achterhoede met naam en toenaam wordt genoemd is kennelijk een positionering buiten het veld van de belanghebbenden (en deskundigen) noodzakelijk. (Vandaar soms de keuze van buitenlandse evaluatieteams bij grote nationale projecten.) De sociale verhoudingen tussen onderzoekers, of het cultureel klimaat verdragen kennelijk niet dat onderzoekers in het openbaar op elkaars werk kritiek uitoefenen. De brief van de 77 oncologen illustreert dit punt ook uitstekend: men weet de uitkomst van de evaluatie al: van het kankeronderzoek kan niets af!" (Hutter 1983*a*, 6)

Rooyen, Doorsma & Eikelenboom 1983), was severely critized as superficial, both in the RAWB report on health research and by Henk Rigter in "Medisch Contact" (Rigter 1984)[102].

Although Rigter and the RAWB had the feeling they had followed a careful course, their critics were not impressed. The Dental Research Section of the Academic Council[103], for example, could not agree less with the methods they used, their judgement on the importance of dental health problems, and the proposal to decrease investment in dental research. Its reaction is typical of this type of debate on research evaluation[104]. The Section explained how dental research capacity had been built up, concluding that one could not speak of stagnation as the RAWB had done[105]. It critized the use of only one reference year and an old one at that (1980 and 1979). The source year 1977 was moreover unjustifiable in the case of dental research because at that time dental research groups were generally small and their productivity fluctuated therefore more strongly than that of big research groups. The section also denied that Excerpta Medica could provide a complete picture of dental research publications. Citing behaviour would moreover be different from that of other specialties because of the small size of research groups. Lastly, the section expressed its amazement that none of the experts known to it had been interviewed by RAWB staff.

Similar comments were produced by others. Some of them stressed the national character of certain types of health research, for example social scientific studies as well as psychiatric and psychological research, making an international criterion like citation frequency in the *SCI* less relevant (Anonymous 1984, 15). A related point of contention was the dominance of medical research in the representation created by the RAWB report, at the expense of "health services research" for example (van der Meer NZI 1984, 2).

On 7 November 1983, the Ministry of Welfare, Public Health and Culture (WVC) invited a number of medical institutions to send commentaries on the RAWB report (van Duyne 1984). The competence of the RAWB was repeatedly brought into disrepute[106]. The National Hospital Institute[107] deemed it "unele-

[102]"Tegen de aanpak die Van Rooyen et al. hebben gekozen, zijn vele gegronde bezwaren in te brengen, onder meer wegens de beperking van het auteurschap tot de eerste auteur; het sommeren van citaties (hetgeen hoogleraren met grote groepen bevoordeelt); het voorbijgaan aan het jaar van publicatie (hetgeen hoogleraren met een lange loopbaan bevoordeelt); het gebrek aan onderscheid tussen publicaties, en het gebrek aan onderscheid tussen soorten disciplines (met een verschillende mate van citatie-rijpheid)." (Rigter 1983*c*, 24)

[103]"Sectie Tandheelkunde van de Academische Raad".

[104]"De Sectie Tandheelkunde heeft met grote zorg en verbazing kennis genomen van het RAWB-rapport en het daarin gestelde over prioriteiten en posterioriteiten bij het geneeskundig onderzoek in Nederland. De uitspraak van de Raad die de haag zijner tanden is ontvloden aangaande het tandheelkundig onderzoek — een verklaring tot posterioriteit — is naar het oordeel van de Sectie zeer aanvechtbaar gezien de kwaliteit van de aangegeven overwegingen." (Dippel 1983)

[105]"van een gebrekkige ontwikkeling is dan ook geen sprake".

[106]For example by the Dutch centre for mental research NcGv: "Van de RAWB had men mogen verwachten dat men zich bij de betrokkenen, in casu vooral het NcGv, grondig zou hebben geörienteerd, alvorens ongerechtvaardigde en schadelijke conclusies de wereld in te sturen." (Anonymous 1984, 14)

[107]"Het nationaal ziekenhuis instituut NZI".

gant" that it had not been consulted (van der Meer NZI 1984, 5). Critics often expressed the feeling that the RAWB did not really appreciate the specific character of their expertise and field of research, especially if they worked in applied or social-scientific research (van Duyne 1984). For example, the National Organization of General Practitioners concluded that the RAWB did not understand the particular position of research on the health system GZO[108] (NHI 1984, 2). Furthermore, it challenged the use of citation analysis and even called a high citation frequency a negative indicator in the realm of GZO[109].

In December 1983, Henk Rigter wrote a new draft background study on priorities in health policy and research (Rigter 1983*d*). The Council did not, however, approve the publication[110]. The RAWB concentrated on its report (RAWB 1983*a*) and tried to ensure that the government followed it up. This was by no means certain: both the medical and the scientific communities were strongly organized and quite capable of torpedoing policies they did not approve. In preparation for a government decision, the RAWB had to write a response to its critics (RAWB 1984) in June 1984[111]. The Council was very critical about the main direction of the debate on its report and warned the government not to give in. It pointed to the asymmetry in the critical appraisal: whereas most of its priorities were not contested, almost of all its posteriorities were, especially by those with vested interests. It expressed its pessimism about the willingness of the medical research community to actively implement priorities[112]. In other words, as long as no research was nominated to be decreased, priorities would not be realized. The RAWB pointed to the diversity of the field, leading to strong differences in the appreciation of the quality criteria and citation analysis. Whereas the cancer researchers critized the RAWB for not relying more on citation analysis, most other

[108]"Gezondheidszorgonderzoek".

[109]"In het RAWB-rapport wordt uitsluitend ingegaan op de wetenschappelijke kwaliteit van het GZO en niet of nauwelijks op de maatschappelijke bruikbaarheid ervan. (...) Daarmee blijft een essentieel wegingscriterium geheel buiten beschouwing. (...) Tegen deze achtergrond is de waardering van de prestaties van het Nederlandse GZO op basis van internationale publicaties, citaties, etcetera op zijn hoogst een criterium van de wetenschappelijke kwaliteit van het onderzoekswerk, maar eerder zelfs een negatieve parameter voor de bruikbaarheid ervan (welke Nederlandse beleidsmaker verdiept zich consequent in "vooraanstaande internationale wetenschappelijke vaktijdschriften?). De prestaties van het Nederlandse GZO zou in ieder geval ook afgemeten moeten worden aan de overdracht van onderzoeksresultaten naar anderen dan vakgenoten" (NHI 1984, 5)

[110]In the RAWB archive a copy of this study carries the handwritten note: "Wim zegt dat dit bij ongeluk in het advies is genoemd, maar dat het niet openbaar is. De Raad wil dat niet."

[111]It also sent replies to its critics separately, e.g. van Bueren (1984). Not all of them were satisfied: the NcGv, NZI, NIPG/TNO and NHI replied to RAWB's reply and insisted that the RAWB report was inadequate (Bensing 1984).

[112]"Over het resultaat van zo'n toetsing is de Raad niet erg optimistisch; weliswaar lijken in het pas gepubliceerde z.g. ontwerp-deelplan Geneeskunde (taakverdelings- en concentratieplan voor het medisch cluster) vrijwel alle RAWB-suggesties tot bescherming gehonoreerd te worden, maar er zijn zoveel andere groei- en beschermde gebieden aan toegevoegd (zulks op andere dan onderzoeksoverwegingen) en er komen maar zo weinig echte krimpgebieden in voor, dat binnen de gestelde randvoorwaarde van een bezuiniging van zo'n f 100 mln het uiteindelijk effect van het deelplan niet anders kan zijn dan krimp over de gehele linie. Van de beoogde en beleden bescherming zal naar de Raad vreest dan niet veel terechtkomen." (RAWB 1984, 2)

critics did not want citation analysis at all. This meant, according to the RAWB, that its decision to adopt not one but five different criteria was "the best possible approach" (RAWB 1984, 4). The council moreover restated its opinion that its use of citation analysis had been more refined, and hence more labour intensive, than the one used by the Foresight Committee Biochemistry. The RAWB stressed the need for an active government policy (RAWB 1984, 5). It took two years for the government to formulate its point of view.

The RAWB did not get all it wanted, but was generally satisfied[113]. Henk Rigter was critical of government policy, stating that government had hardly done anything itself and could therefore not expect one report to solve all the problems[114]. His irritation was apparently triggered by the cautious way the government trod with regard to the sharp criticism from the scientific community. It endorsed the main thrust of the RAWB report[115] (Anonymous 1985), but also acknowledged that the critics had a point. The cabinet presented its own criteria for assessing the function of the healthcare system and used these to partly distance itself from the RAWB report[116]. The government did not think the three-way schedule proposed by the RAWB was feasible[117]. The relationship between fundamental biological research and the healthcare system would be "more complex and indirect" (OenW 1985, 5) than the RAWB report suggested. The same cautious attitude was adopted with respect to the methods used to assess the quality of research: although critical remarks were possible, the overall RAWB judgement was underwritten[118].

Remarkably, the government statement explicitly devoted a few sections to citation analysis, citing it as a valuable new tool for science policy although indi-

[113]The feeling was aptly formulated by Henk Rigter, who had become secretary of the new Health Council, in a letter to his former colleague Wim Hutter, secretary of the RAWB: "Mijn algemene indruk bij het lezen van het stuk was: laten we ons niet druk maken; we krijgen materieel op alle punten gelijk — de rest is een ambtelijk achterhoedegevecht. Niettemin schemert het mij iedere keer weer voor de ogen (ik leer dat maar niet af) als ik het wereldvreemde commentaar lees van onze beste stuurlui op de wal. Hoe komt men erbij om bij het ongelofelijk vele werk dat we verricht hebben, ook nog eens te suggereren dat we concrete mogelijkheden van internationale taakverdeling en samenwerking in onze beschouwing hadden moeten betrekken? Hadden we onze analyses dan tot de hele wereld moeten uitstrekken?" (Rigter 1985)

[114]"Waar het beleid op deze punten zelf weinig tot niets gepresteerd heeft, is het naïef te verwachten dat dat in één enkel advies rechtgetrokken kan worden".

[115]The official publication in "De Staatscourant" carried the headline "Kabinet grotendeels eens met RAWB-advies 'Prioriteiten gezondheidszorg"'

[116]"De regering is derhalve van mening dat de overwegingen die de Raad laat gelden bij het stellen van prioriteiten, te weten de kwaliteit en omvang van het onderzoek en de relevantie ervan voor het volksgezondheidsbeleid op zich onverkort van toepassing zijn, maar dat een uitbreiding tot een ruimer kader nodig is door ook de bovengenoemde functies van het onderzoek een rol te laten spelen bij de te maken keuzes." (OenW 1985, 3)

[117]"De uitvoerbaarheid van dit drietrapsproces wordt door de regering echter met enige scepsis bezien, met name ook gegeven de wetenschappelijke, maatschappelijke en politieke realiteit" (OenW 1985, 5).

[118]"De analysemethoden van de Raad bevatten vele waardevolle elementen, hoewel bij onderdelen van de methodiek kanttekeningen kunnen worden gemaakt. Desondanks is de regering van mening dat het kwaliteitsbeeld dat de Raad geeft van het gezondheidsonderzoek in het algemeen kan worden onderschreven." (OenW 1985, 6)

cators could not directly measure scientific quality[119]. The main objections mentioned by the government in its statement were its limited value if journal publications were not the predominant form of output, its limitations in asessing national research, and its focus on past performance. Interviews with foreign and national experts did not solve the problem, because "subjective plus objective does not equal more objective". The government moreover questioned the independence of the various measures used by the RAWB and thereby the extra value added by a more refined citation analysis. According to the Dutch government, the RAWB report was therefore less firmly entrenched in reality than its authors claimed[120].

6.6.3 RAWB's indicator policy

The medical project was an important part of RAWB's indicator policy and helped in shaping it. As the Council stated in a draft note at the beginning of its fourth term, it wished to be a well-informed centre of science policy and would therefore produce science and technology indicators if necessary[121].

In the first half of the 1980s, the Council discussed the potential and the pitfalls of quantitative science and technology indicators on several occasions. Advice to the government about the use of indicators by other agencies was combined with its medical project and the preparation of a new series of Indicator Reports it planned to publish from 1983 onwards. One such occasion was its official commentary on the report of the Evaluation Committee Biochemistry in 1983. In November 1982, this committee had published its report "On life"[122] (Biochemie 1982). It was the first time that citation analysis was used in a Dutch disciplinary evalution report (Van der Meulen 1992, 31–32) and it provoked a lot of discussion. The committee ranked biochemistry research teams based on citation analysis using the *SCI*. It explicitly rejected measuring quality by counting the number of publications[123]. The biochemistry committee built on the evaluation work of physics at FOM which it considered to be "pioneering". It

[119] "De belangrijkste maatstaf voor de Raad om tot een waardering van de kwaliteit van het onderzoek op vakgroepsniveau te komen is de citatieanalyse, en wel in vier verschillende vormen. Hierdoor is het voor het eerst mogelijk een indruk te krijgen van de prestaties van het gezondheidsonderzoek in Nederland. (...) In haar reactie op het rapport van de Verkenningscommissie Biochemie heeft de Regering aangegeven dat zij wetenschapsindicatoren een nuttig instrument vindt om uitstraling naar en gevolgen van wetenschappelijke publikaties op de onderzoekwereld aan de orde te stellen, maar dat het daarmee nog niet een directe maatstaf is voor gebruik, laat staan voor kwaliteit." (OenW 1985, 7)

[120] "Hoewel het beeld dat het RAWB-advies geeft van de kwaliteit van het onderzoek dus wellicht geen vijfvoudige verankering heeft, meent de regering dat het beeld in globale zin moet worden onderschreven en niet afhankelijk is van details van de methodiek."

[121] "De RAWB behoort een goed geïnformeerd centrum met betrekking tot het "wetenschapsbedrijf" te zijn (blijven). Het analyseren en zo nodig produceren van science & technology indicators is daarom een vaste taak van de Raad." (RAWB 1982*a*, 2)

[122] "Over leven"

[123] "De commissie wil dan ook waarschuwen tegen de tendens om het aantal publikaties zonder meer te hanteren als criterium voor de kwaliteit van wetenschappelijk onderzoek." (Biochemie 1982, 78)

considered most objections to citation analysis to be either invalid or irrelevant (Biochemie 1982, 80), with the exception of the problem of differing citation patterns in different specialties. The committee tried to solve this problem by subtracting self-citations from the scores and by normalizing this number of citations for each specialty within biochemistry. Given this, it concluded that a high citation score correlated with a high quality. The committee also emphasized that qualitative indicators implicitly used in the peer review procedures also had many shortcomings (Biochemie 1982, 262–263). Although the committee used five different indicators, the main one was the corrected number of citations per publication and per researcher (Biochemie 1982, 77–99,257–269). It also looked at the groups with extraordinarily high citation scores.

The discussion about the use of citation analysis triggered by the biochemistry report, was aggrevated by a miscalculation due to which a research group in Leiden felt itself disadvantaged[124] and by the problem of misspellings which resulted in an average underestimation of the citation score by 45 percent. The Minister of Education and Science Policy subsequently asked the RAWB to explicitly discuss the methodology used by the biochemistry committee in its report (Hutter 1983*b*, 5). In its discussion, the RAWB compared its own experiences in the medical project with those of the Evaluation Committee. The biochemists relied more heavily on ISI-made citation analysis than Henk Rigter, who had collected the citation data by hand. The Council expressed its amazement that the biochemistry committee had not used its own expertise as a check on the bibliometric data. The report strengthened the Council in its position that citation analysis should always be combined with expert knowledge of the field[125]. It recommended that subsequent evaluation committees would use the instrument only in a comparative way. In a note to the Council, secretary Wim Hutter moreover warned against the proliferation of citation analysis as a stand-alone product.

In the same year 1983, the Council started its own Science and Technology Indicators (WTI) Project with the ambitious aim of producing a quantitative description of all research in the Netherlands. Preparations had started the year before. Staff member Arie van Heeringen travelled to the US to gain insight into the way NSF prepared its Science Indicators Reports[126]. Van Heeringen concluded that a science indicators report was indeed useful, and recommended that as many indicators as possible should be used since each indicator had its shortcomings[127]. A science indicators report should moreover not restrict itself to quanti-

[124] The ISI files missed a highly cited article from this group.

[125] "De conclusies van de Raad zijn nu de volgende: citatie-analyses, door deskundigen uitgevoerd en toegepast, zijn voor de hoofdzakelijk op kennisvermeerdering gerichte exacte en levenswetenschappen een nuttig instrument om zicht te krijgen op de kwaliteit van onderzoekgroepen, vooral in vergelijkende zin." (Hutter 1983*b*, 8)

[126] "Het voornaamste doel van mijn bezoek aan de Verenigde Staten was een inzicht te krijgen in de totstandkoming en het nut van de Science Indicators-rapporten zoals die iedere twee jaar in de Verenigde Staten verschijnen. Dit mede met het oog op de mogelijkheid ook voor Nederland een dergelijk rapport samen te (laten) stellen." (van Heeringen 1982, 3)

[127] "Een analyse van de situatie van het onderzoek in Nederland dient met zoveel mogelijk interpreteerbare indicatoren te worden gestut. Elke indicator heeft wel zijn tekortkomingen, de perfecte indicator bestaat niet; de stelling dat indien meerdere indicatoren een zelfde beeld opleveren,

tative data but should also include qualitative interpretations[128]. Given the state of flux science indicators research still was in, Van Heeringen did not think it was wise to give any one government department the responsibility for writing the indicators report. He therefore advised the RAWB to take the responsibility itself given its position between policy and research[129].

The WTI project aimed at producing a general quantitative description of research in the Netherlands, using both input and output indicators[130]. The bibliometric output indicators would mainly be used to compare the position of Dutch research and development with other countries. The humanities and social sciences would merely be described by available ISI data. The natural and technical sciences were supposed to be analyzed in a more sophisticated way. Apart from bibliometric description, a cluster analysis of the *SCI* would be used to perform an analysis of the strength and weakness of Dutch research[131]. The citation frequency was the main indicator used in the natural sciences. With regard to the technical sciences the RAWB staff wished to use patent data, data of licences and royalties, and data of the international trade in knowledge intensive commodities.

The Indicator Report was published in June 1984 and compared Dutch research in the preceding year with that of its foreign competitors (van Heeringen et al. 1984). It was the first Dutch report to use co-citation analysis to map the scientific output of a country and to reveal the underlying structure of science[132].

---

dit beeld een meer algemene betekenis heeft, is mijns inziens serieus genoeg om een SI-rapport op te stellen en deze in discussie te brengen" (van Heeringen 1982, 12).

[128] "de gegevens dienen te worden geanalyseerd en geïnterpreteerd op basis van verricht onderzoek naar de waarde en de beperkingen van de betreffende indicatoren" (van Heeringen 1982, 12).

[129] "De RAWB lijkt een geschikte kandidaat om dit karretje te trekken; ze staat in bepaald opzicht tussen het beleid en de onderzoekers in en heeft goede contacten met beide. Daarnaast kan door deze koppeling met de RAWB de signaalfunctie die een SI-rapport kan vervullen goed tot haar recht komen." (van Heeringen 1982, 13)

[130] "Het Wetenschaps- en technologie-indicatoren project beoogt een zoveel mogelijk kwantitatieve beschrijving te geven van het Speur- en Ontwikkelingswerk (S & O) in Nederland. (...) Wij denken de Nederlandse S&O te karakteriseren met behulp van zogeheten wetenschaps- en technologie-indicatoren, die doorgaans verdeeld worden in "input" en "output" indicatoren. Inputindicatoren beschrijven financiële en personele middelen e.d., outputindicatoren zaken als patenten, bibliometrische gegevens etc.. Nu heeft elke indicator wel zijn specifieke 'tekortkomingen' die de interpretatie bemoeilijken; de indicatoren geven derhalve maar een gebrekking inzicht in de vragen waarin wij zijn geïnteresseerd. De specieke kracht van een WTI-rapport, zoals wij dat voor ogen hebben, is gelegen in het feit dat verschillende indicatoren tegelijkertijd in de analyse worden betrokken zodat een beter inzicht in het totaal kan worden verkregen." (Mombers, van Venetië & van Heeringen 1983, 1)

[131] "Naast deze min of meer gebruikelijke analyse van publicaties zal een sterkte-zwakte analyse van Nederlandse onderzoek in vergelijking met het buitenland (de clusteranalyse van de Science Citation Index) zeer waardevolle gegevens kunnen opleveren. Met deze techniek kan een beeld geschetst wroden van: 1. de mate waarin Nederland in een bepaald onderzoeksgebied is vertegenwoordigd; 2. de "impact" van Nederlands onderzoek in een bepaald veld, vastgesteld aan de hand van citatiefrequenties naar Nederlandse publikaties; 3. belangrijke buitenlandse onderzoekspartners, bepaald door de citatiefrequenties van Nederlandse onderzoekers naar buitenlandse collegae. 4. de belangrijke landen in een bepaald onderzoeksveld." (Mombers et al. 1983, 4)

[132] "Met de clusteranalyse beschouwt men niet alleen de output, maar ook de gehele structuur

Co-citation analysis was presented as a more objective method of distinguishing between specialties than either the traditional distinction between natural science, social science, and the humanities or the somewhat arbitrary assignment of journals to certain areas. After all, authors assign themselves to certain clusters, which would mean that co-citation clustering was a more "objective method", the report explained (van Heeringen et al. 1984, 51). It enabled an easy specialty-by-specialty comparison between Dutch and foreign scientific acitivity and impact. Comparing clusters in different years would moreover enable an analysis of cluster dynamics, including the identification of emerging research areas. Last but not least, this method could produce maps of science since co-citation links existed by definition not only within clusters but also between them. The report contained several examples of these co-citation maps of Dutch science (van Heeringen et al. 1984, 67–86).

RAWB's first indicator report (van Heeringen et al. 1984) was followed up by a second indicator report in 1988 (van Heeringen & Langendorff 1988) in which, incidentally, no co-citation analysis was applied. Since then, the production of this type of report has been taken over by a specialized indicator study center, founded at the University of Leiden with RAWB support.

6.6.4 Ministerial support for an observatory

In the second half of the 1980s, the Dutch Ministry of Education developed an ever-growing appetite for information regarding the scientific system. According to the documents circulating at the ministry at the time, this was related to a shift towards a "more analytical" type of science policy. To enable this, the quality of information used in policy documents should be improved, and the ministry cooperated with the indicator specialists ("the indicator club[133]") at the University of Leiden (OenW 1987*c*). As a follow-up on the RAWB indicator report 1983, a Leiden Science and Technology Indicator Project had been set up headed by the physicist Ton van Raan. Van Raan had previously worked at FOM where he had acquainted himself with scientometrics under the guidance of Cees le Pair. This work on indicators was organized in LISBON, which had the status of a faculty working group but functioned as a separate institute, according to its leader Ton van Raan[134]. Several projects were conducted, exploring the various possibilities of scientometric analysis. The main focus was on the possibilities of mapping science with co-citation clustering, the measurement of the humanities and social sciences, the scientific base of technological developments, the use of indicators in applied science, and early warning indicators (van Raan 1987*c*). For the ministry, this work had to result in a policy-oriented information system. Ton van Raan wished to do a pilot study first, focusing on agriculture and the development of electronics (van Raan 1987*b*).

die daar achter ligt. Zo'n analyse geeft dus een veel completer beeld van het wetenschappelijk bedrijf dan een indicator die een gebied slechts voor een deel of op een indirecte manier beschrijft." (van Heeringen et al. 1984, 88)

[133]"De indicatorenclub".

[134]Ton van Raan, Interview, 1995 Leiden.

In 1986, a new information project within the Ministry of Education was started, "R&D in the Non-Governmental Sector"[135] (SONO) to get insight into the quantitative and qualitative trends in research in the private sector (SONO 1985, 1). Attention shifted from indicators on developments in Dutch research to developments within problem areas of government policy (Langendorff 1986). To enable this, a restructuring of information underlying the two-yearly Science Budget was considered with the aim of refining and increasing available information (OenW 1986*a*, OenW 1986*b*). The officials were setting up a fully automated information service, including output data. One of the ideas was to co-operate with a limited number of countries (Germany, UK, France, Sweden and/or Belgium) (OenW 1986*a*, OenW 1987*d*, 26) in developing an advanced information system, using the US Science Indicators Reports as an example. The indicator reports of the European Union and the OECD were not completely satisfactory[136]. In the Science Budget of 1987, more quantitative information than ever before was presented. Yet, the science policy officials at the Ministry were not satisfied: they wanted more quantitative information as well as more in-depth analysis of these data (OenW 1986*c*). This increasing ministerial appetite for information and indicators was seen as related to the trend to "govern at a distance" with its decreasing importance of directives (Dits 1987*b*, 5).

In June 1987, a working group on indicators was formed within the Ministry to speed up indicator development. Apart from refined information to be included in the Science Budget, the focus was on evaluation of research institutes, identifying the strong and weak points of Dutch science and, last but not least, early warning indicators (Dits 1987*b*, 2). The working group was to remedy the perceived lack of a consistent strategy with regard to indicators (Dits 1987*a*). For example, the focus was directed too much at "fancy" indicators and not enough at the effective use of the more common ones and the building up of a database. The Ministry reached an agreement with the Leiden group LISBON in the form of the project "Trends in Dutch research"[137] which focused on instrumentation, demographic developments in science, and international cooperation by scientists (van Raan 1987*a*)[138]. The results were to lead to "convincing materials that can be used for political decisions"[139] (OenW 1987*e*). LISBON proposed to start with a pilot study of agriculture and technological developments in electronics (van Raan 1987*a*, 2) to find out which indicators and data would be suitable.

According to one of the ministry officials, Henk Dits, a choice had to be made between two possible indicator models (Dits 1987*a*). The first was one in which a complete review of science was made every few years. Indicators had to enable comparison between the actual state of affairs and policy goals. An indicator

[135]"Speur en Ontwikkelingswerk in de Niet-Overheid sector".

[136]"Gevoelen is dat EG en OECD weliswaar nuttig werk doen, maar de vanuit deze organen geleverde informatie wordt op een aantal punten onvoldoende geacht." (OenW 1986*d*, bijlage 4)

[137]"Trends in het Nederlandse Speur- en Ontwikkelingswerk".

[138]A first meeting on this project took place on 16 June 1987, when representatives of LISBON, the ministry of education, and the central Dutch statistics institution "het Centraal Bureau voor de Statistiek" CBS met. (OenW 1987*a*, OenW 1987*e*, van Raan 1987*a*).

[139]"Voorts is hier van belang dat de analyse resultaten moet opleveren die als overtuigend materiaal kunnen dienen bij politieke beslissingen." (OenW 1987*e*)

strategy in this model had to strengthen the primary information service (providing data for the indicator values) as well as to build new indicators [140]. This model aimed primarily at national state level. Research organizations would have to build their own indicators. The second model, on the contrary, departed from the wish to strengthen the independance of research institutions and intermediary organizations. Indicators would therefore be primarily at the service of these organizations, with the ministry as the stimulating agent. The function of indicators in developing science policy would, though still important, be secondary. The Action Group Indicators preferred the second model combined with the primary information system from the first model (OenW 1987*b*). Indicators should moreover be made compatible at the international level (EC, OECD).

The work of the action group indicators resulted in an "Indicators Plan of Work". The document was based on the assumption that the use of indicators in science policy would increase, "in any case as a means to formulate and underpin policy"[141] (OenW 1987*f*). According to the plan, three problems were paramount: the policy questions were often insufficiently articulated; the combination of necessary indicators was unclear; and the primary data were often incomplete or lacking. The first phase was therefore to be exploratory. The plan proposed giving LISBON the assignment of interviewing international experts. At the same time, the Action Group would conduct an inventarisation within the ministry to assess their information needs. The subsequent combination of "demand" and "supply" was to lead to a draft policy plan for indicators. This plan would be discussed at an international workshop in April 1988, leading to a definitive policy plan and the selection of pilot indicator studies at the end of 1988[142].

This ministerial activity laid the foundation for the Netherlands Observatory of Science and Technology (modelled on the French example, "l'Observatoire des Science et Technique") which since 1994 has become the main national producer of science and technology indicators for science policy. The regular production of these highly visible indicator reports, seem to further encourage the use of indicators in lower-level assessment exercises by universities (cooperating in the Cooperation of Dutch Universities VSNU) and research councils. Indeed, it is now official policy to use bibliometric analysis in research evaluation exercises within the context of qualitative assessment of research performance in all of the sciences, social science and humanities (Wouters 1998*a*).

[140] "De strategie (voor de werkgroep indicatoren) wordt dan het uitwerken en uitvoeren van een programma dat enerzijds gericht is op het versterken van primaire informatievoorziening, om voldoende en geschikte gegevens voor reeds bestaande indicatoren te verkrijgen. Anderzijds gericht op het totale overheidsfunctioneren t.a.v. onderzoek en ontwikkeling." (Dits 1987*a*, 2).

[141] "De werkgroep is uitgegaan van de verwachting dat het gebruik van indicatoren in het wetenschapsbeleid zal toenemen, in ieder geval als een hulpmiddel bij het formuleren en onderbouwen van beleid" (OenW 1987*f*, 1).

[142] A supervising committee was formed by the ministry, LISBON and the RAWB. A special wish of the Science Policy Department was stressed as well: early warning indicators. "Drie ton van de WB begroting via RAWB in het programma geinvesteerd. Van de drie ton mogen we toch minstens 'early warning' indicatoren op enkele proefgebieden verwachten." (OenW 1987*f*, 5)

Back to Contents Page

Back to Contents Page

Chapter 7

Scientometrics

7.1 Introduction

In 1978, the journal *Scientometrics* was launched as a new medium to stimulate the development of scientometrics, the quantitative study of science. According to one of the editors, it would be "a journal for the publication of meaningful and valuable contributions to this new field of science" (Beck 1978). Derek de Solla Price, one of the editors-in-chief, described the development of scientometrics as the emergence of a "relatively hard" social science. "For many years now", he wrote

> we have been guest editors in the journals of other neighboring fields and the special bibliographies in bibliometrics and science of science testify to the rapid cumulation of a coherent literature. (...) I use this word to imply that the growth is coherent, with the *new* advances being laid down on the basis of rather fresh preceding foundations for the new growth. Thus, the relatively hard sciences are distinguished from those that are relatively soft. (Price 1978)

It was Price's goal to develop scientometrics as a "hard social science":

> I have to believe that if the little green alien people came from a distant planet and communicated with us, all else about them might be alien but they would know in some fashion or other such things as Planck's Constant, the velocity of light and the Wave Equation. I believe they might also find reasonable points of correspondence with our scientometrics even if their social arrangements were utterly different from our own.

It is a known fact that the creation of a scientific journal is a further stage in the development of emerging specialties. It creates a publication outlet for the specific worldview the participants in the new endeavour share. The creation of the journal *Scientometrics* likewise signaled a new phase in the development of scientometrics. Tibor Braun, an analytical chemist working at the Hungarian Academy of Science Library's information service since 1975[1], took the initiative

[1] Tibor Braun, Interview, Budapest, 29 July 1994.

and founded the journal in 1978. He was a new player in the field and most scientometricians, including Eugene Garfield and his colleagues at ISI, were taken by surprise. They had not expected that someone would create a journal specifically devoted to scientometrics and the science of science so soon. Braun used his previous experience in publishing. In 1968 he had created an international journal in radioactive chemistry, followed by two more, published by Elsevier Science. In the mid-seventies, Braun encountered the works of V. V. Nalimov, Derek Price, Michael Moravcsik and Eugene Garfield. He organized a seminar at his chemistry institute, focusing on the evaluation of Hungarian research by scientometric means, which proved to be the starting point of the journal.

From the very beginning, the journal has leaned towards the "science of science" approach. "I consider scientometrics as a hard social science", Braun stated in 1994[2]. It is true that other journals also publish scientometric research. Most of these are located in science studies, in library information science and in the science policy domain. Nevertheless, by any standards *Scientometrics* still is widely regarded to be scientometrics' core journal. For example, it is the only one regularly publishing papers on data collection and scientometric methods. This chapter[3] therefore presents a scientometric self-portrait of scientometrics' core journal as a means to acquire more insight into the culture of this specialty.

7.2 Collection and organization of the data

The complete bibliographic description of all articles and notes in the first 25 volumes of *Scientometrics* (1978-1992) from SCISEARCH/DIALOG, were downloaded. The bibliographic description includes: the names and addresses of every author; the year of publication; the title, volume number and page numbers of the publication; the number of references; and for every reference the name of the first author, the journal title, the volume number, the year and the page numbers of the journal cited. Wouters & Leydesdorff (1994) organized these data in relational database files (dBase III^{+}). These data were analyzed by programming in dBase III^{+} on a 386 SX personal computer (16 Mhz, 6Mb RAM), running DrDos 6.0 as operating system. Because dBase is a line based programming environment, this produced huge computer programs. The organization of the data, while adequate for its task, was relatively bulky, due to restrictions in the software and the disk operating systems. Reproducing the analysis after several years was therefore not without difficulty.

Instead of trying to re-use the programs that had been written at the time, it was decided to start all over again in order to write this chapter. The same raw datafile was used. But none of the data correction programs, databasefiles or data processing programs was used. This re-analysis was performed on a Pentium personal computer (133 MHz, 64 Mb RAM) running Linux 2.0.30 (Red Hat

[2]Tibor Braun, Interview, Budapest, 29 July 1994.

[3]I would like to thank Dr. Loet Leydesdorff for his generous guidance in the research underpinning this chapter and for his permission to use this in this thesis. This text is a partly rewritten version of a joint article (Wouters & Leydesdorff 1994).

4.2) with X-Windows (XFree86 with AcceleratedX). The datafile was analyzed with PERL (the Practical Extraction and Report Language)[4]. Running a particular program now took minutes, instead of the hours and hours, and in some cases even the days and days previous programs had been taking. Because of their text-oriented nature, Perl (Katz 1995) and Unix provide a very precise way of controlling regular expressions, thereby facilitating managing the irregularities in the data[5]. The problems of varying ways of spelling names[6] has been solved by identifying authors, institutions and journals with unique ID-numbers.

7.3 General features

Since 1978, 779 items have been published in *Scientometrics*. They contain 11285 valid references[7]. The number of publications per year increases linearly with time (figure 7.1 on page 170). After a steep growth in the first three years, the number of publications increases with on average 3.5 publications per year.

The data set behaves in many ways like the average bibliometric data collection. Skewed distributions of properties over units of analysis dominate. Plotted on a bi-logarithmic scale, the relationship between number of authors and the number of publications of these authors is an almost perfect straight line (figure 7.2 on page 170). This means that most authors have published only once in *Scientometrics*, whereas a few are very productive.

The same holds for the institutions that have published in *Scientometrics* figure 7.3 on page 171. By far most institutions have only published once. The distribution of citations over authors is likewise skewed figure 7.4 on page 171.

[4]This programming language is also called the Pathologically Eclectic Rubbish Lister. It is a flexible and powerful programming environment for analyzing textual data (Perl combines the features of C with the versatility of UNIX shell programming). (See the appendix for a list of the PERL programs I wrote and used).

[5]Therefore, error correction could be done on the fly: it was no longer necessary to create "cleaned up" datafiles with their potential sources of additional errors. Every program started with the same datafile, thereby ensuring a greater consistency of the analysis. The following types of irregularities were found: misplaced control characters, missing "end of record" characters, missing "end of field" characters, unexpected strings of spaces or newlines, typos, empty references, references to non-existent publications, various ways to write one name, and various ways to write one institutional address.

[6]One of the extreme examples is Derek de Solla Price, whose name has been recorded as: DESOLLA D, DESOLLA PD, DESOLLA PDJ, PRICE D, PRICE DD, PRICE DDS, PRICE DJ, PRICE DJD, and PRICE DS.

[7]References to a year before 1200 AD, references to years after the year of the citing publication and empty references were excluded.

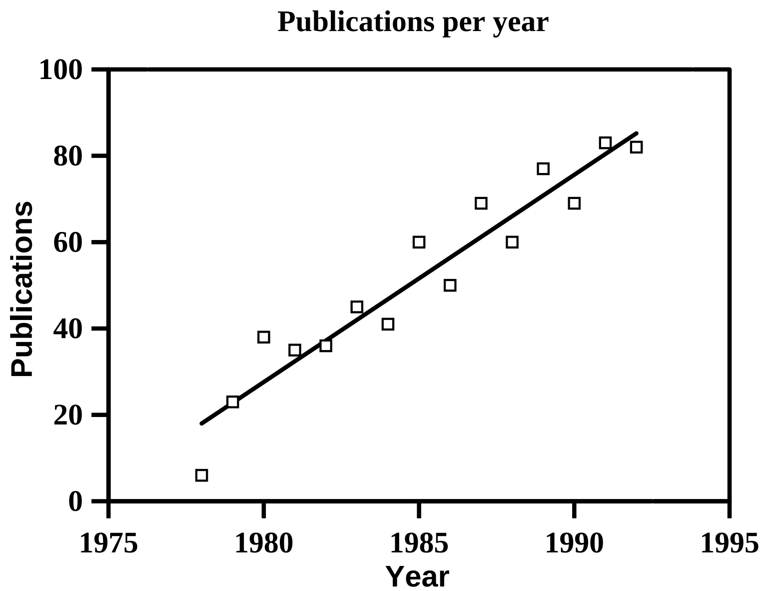

Figure 7.1: The number of publications per year.

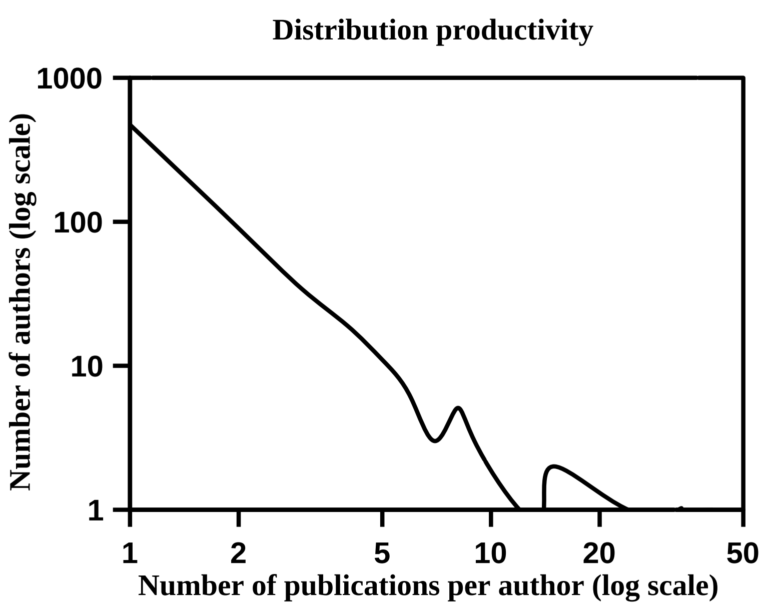

Figure 7.2: The number of publications per author in relation to the number of authors.

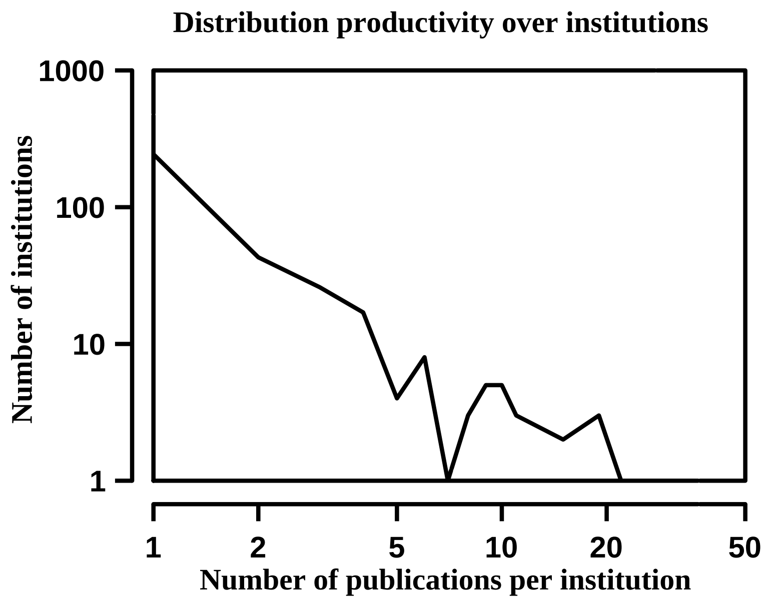

Figure 7.3: The number of publications per institution in relation to the number of institutions.

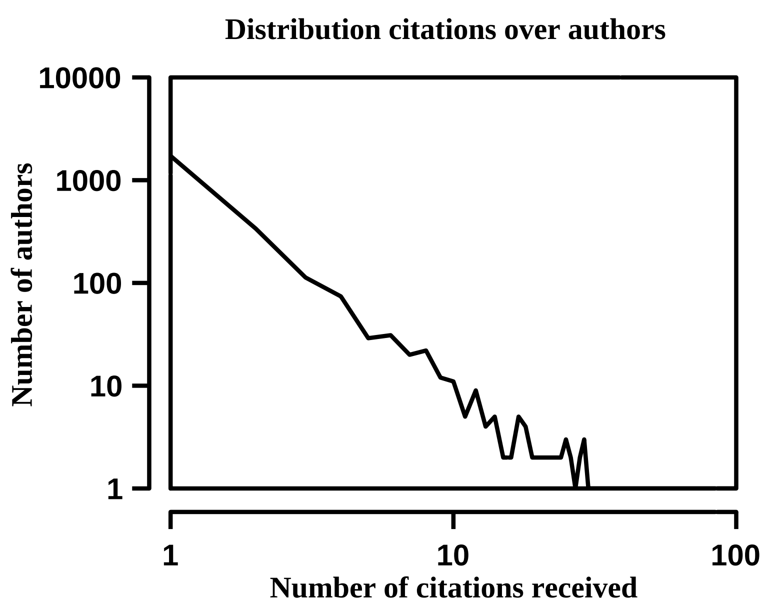

Figure 7.4: The number of citations in relation to the number of cited authors.

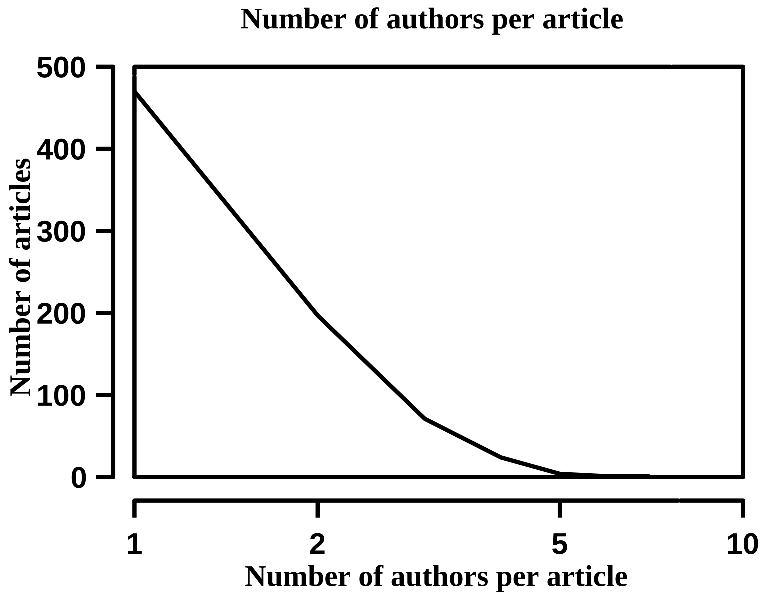

Figure 7.5: The number of authors in relation to the number of articles.

The publications have been written by 669 different authors, affiliated to 389 different institutions. Cited and citing authors together total 4441 unique names. On average, every author has published 1.8 times and every paper has been written by 1.6 authors. Most articles have been written by one author, the maximum number of authors being seven (one paper) (figure 7.5 on page 172). The distribution of this number of authors over the articles is less skewed then the number of citations: a straight line is already approached in a uni-logarithmic plot. The skewness of the distribution of the number of institutions that have jointly published over the total number of institutions has an intermediate skewness: it is less skewed than the distribution of citations, but more so than the distribution of co-authors figure 7.6 on page 173. Most articles were published by one institution.

The 779 publications cited 3971 different journals over 25 years. Most of these journals were cited only once: 2930. These recurring skewed distributions are a normal bibliometric feature. In this sense, the data from *Scientometrics* confirm the non-normal distribution of bibliometric properties.

7.4 Has Price's dream come true?

7.4.1 Method

According to Price's theory of knowledge growth (Price 1965*a*), science distinguishes itself from other fields of study by the way scientists refer to their literature (Price 1970). The existence of "research fronts" in science supposedly leads

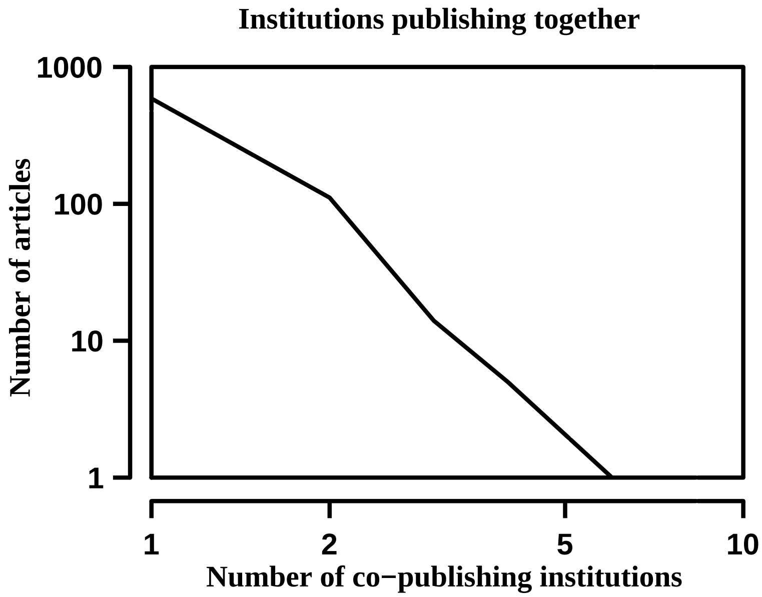

Figure 7.6: The number of co-publishing institutions in relation to the number of articles.

to an "immediacy effect", which can be measured in terms of the so-called *Price Index*. The Price Index is defined as the proportion of the references to literature over the last five years. Price estimated that this index would vary between 22 and 39 percent if no immediacy effect were present. A field that is all research front and with no general archive might have a Price Index of 75 to 80 percent. From his analysis of 162 journals, Price (1970) concluded:

> Perhaps the most important finding I have to offer is that the hierarchy of Price's Index seems to correspond very well with what we intuit as hard science, soft science, and nonscience as we descend the scale.

Biochemistry and physics are at the top, with indexes of 60 to 70 percent, the social sciences cluster around 42 percent, and the humanities fall in the range of 10 to 30 percent. Cozzens (1985) corroborated with Price in his observations. Marton (1985) also supported Price's immediacy factor. Moed (1989) found that the overall picture may be a bit more complicated than Price thought: within the natural sciences, significant differences in the Price Index may occur. This author suggested that high Price Indices correlate with high citation scores.

Price determined the Price Index by taking all references in a given year and counting the number of references to literature published in the preceding five years. Moed (1989) proposed the use of the Price Index of the references of every citing article as a basis for computing the Index for a journal (or a specialty) as a whole. Moed's method has the citing article as the unit of analysis, enabling the analysis of the distribution of the Price Index over the set of articles. This study analyzed the Price Index in both ways. For each citing article the number of articles cited less than six years old, was counted as a fraction of the total number of articles cited by that particular citing article. In other words:

$$PI = \frac{N_1}{N_2} \times 100 \tag{7.1}$$

where N_1 is the sum of cited articles for which holds:

$$PY_{citing} - PY_{cited} \leq 6 \tag{7.2}$$

and N_2 is the total number of references of the citing article. The Price Index of *Scientometrics* is then simply the average of the Price Index over all articles:

$$PI = \frac{\sum pi}{N} \tag{7.3}$$

where N is the total number of articles published (779). The distribution of the values of the Price Index over the total set of citing articles was also computed.

The Price Index, computed according to Price's method, takes the citing year as the unit of analysis. For all articles published in a given year, the number of references less than six years old, were counted as a fraction of the total number of references (7.1). This gives the Price Index for every publishing year. The Price Index of *Scientometrics* is then the average of the Price Index over all years: (7.3), where N is the total number of years (25).

7.4.2 Results

The average Price Index of *Scientometrics* measured by Price's method is 41.73 percent. Counted according to Moed's method its value is 47.76 percent[8]. Over the years, the value of the Price Index seems to stabilize figure 7.7 on page 175. The distribution of the Price Index over the articles figure 7.8 on page 176, is a superposition of three different distributions: one subgroup (29 articles) has an index of 0, a second subgroup (128 articles) has an index of 100 and the third subgroup is a irregular normal distribution of around 50 percent. Correcting the computation of the Price Index by discarding the two subgroups with a Price Index of 0 and 100 does not change the development of the Price Index over time. Thus, the Price Index remains stable over the years. The Price Index displays neither rise nor fall from 1978 until 1993. This means that scientometrics has not become "harder" in this period[9]

[8]These values differ slightly from those found by (Wouters & Leydesdorff 1994). This is because the previous analysis excluded all references to years after 1992 whereas in the present analysis all references were excluded to articles "younger" than the citing article. The present measurement is more accurate.

[9]Recently, Schubert & Maczelka (1993) concluded from an analysis of *Scientometrics* in 1980–81 and 1990–91 that the journal has moved slightly from the "soft" (social) towards the "harder" (natural) sciences. They drew this conclusion from the rise of the Price Index from 35 percent to 42 percent between these measurement points. This observation is, however, based on only two measurement points. Because of the statistical fluctuations in the value of the Price Index over time, this may not be a sufficient number of measurements. The present conclusion may therefore be seen as a proposal to correct the one Schubert & Maczelka (1993) drew.

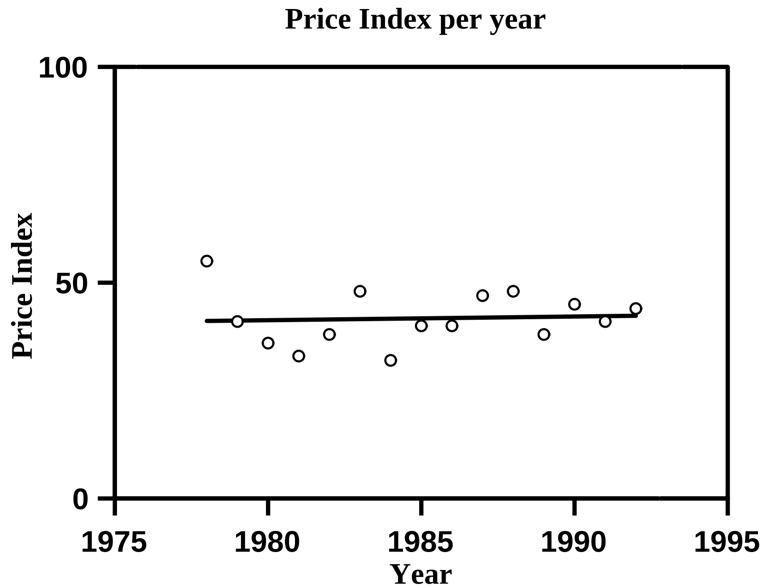

Figure 7.7: The value of the Price Index per year, Price's method

How does this value of the Price Index compare with neighbouring journals? Spiegel-Rösing (1977) found that the Price Index of *Science Studies* (vol. 1-4) is 41 percent. To get a more comprehensive picture, the Price Index of *Scientometrics*, *Social Studies of Science* and the *Journal of the American Society for Information Science* were compared with one another from 1979 up to and including 1992. *Jasis*'s average Price Index is 45.8, *Social Studies of Science* has an average index of 36.8. On a Tukey Test the differences between *Social Studies of Science* on the one hand and *Jasis* and *Scientometrics* on the other are significant at the 5 percent level. The difference between *Scientometrics* and *Jasis* is not significant. These values, however, are all within the range Price indicated for the social sciences in general. These findings are consistent with (Moed 1989) who found significant differences between Price Indices of subfields from the same discipline.

In accordance with Price's theory, the number of references to literature of a specific age rises until the cited literature is two years older than the citing literature, and then falls off figure 7.9 on page 176. This decline is, however, very gradual. There is no clearly visible "immediacy effect". Apparently, scientometrics does not have a clearly defined research front. Old publications still count. Within the framework of Price's theory, this means that scientometrics does not behave like the paradigmatic natural science in which supposedly only recent publications are cited.

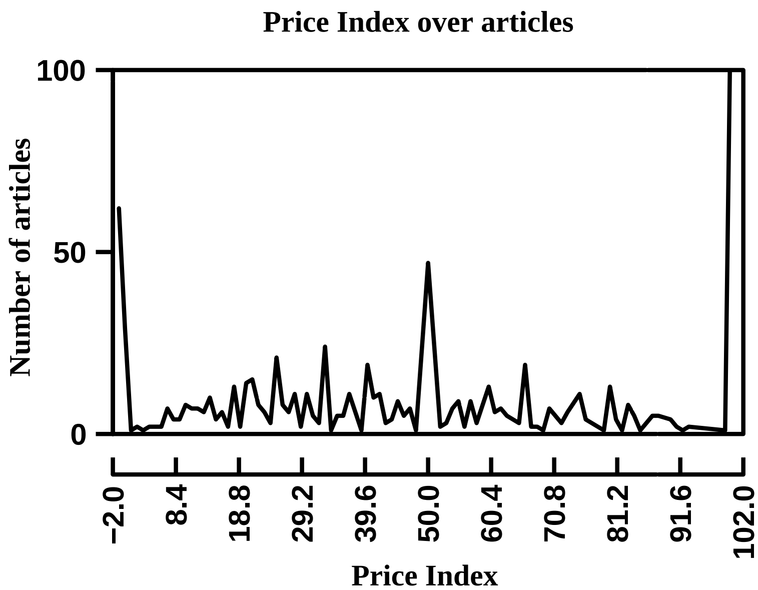

Figure 7.8: The distribution of the Price Index over the articles, Moed's method

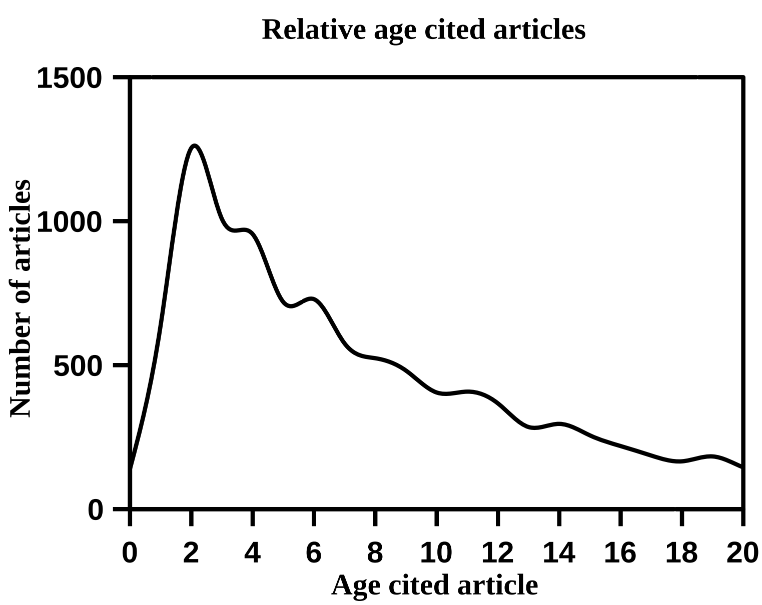

Figure 7.9: The age of cited articles relative to their citing articles in relation to the number of cited articles

Number of publications	Author
41	SCHUBERT A
34	MORAVCSIK MI
33	BRAUN T
24	GLAENZEL W
15	HAITUN SD,LEYDESDORFF L
14	VANRAAN A
12	COURTIAL JP
9	MARTIN BR,NARIN F TODOROV R
8	NEDERHOF AJ,ROUSSEAU R SENGUPTA IN,SMALL H VLACHY J
7	LANCASTER FW,VINKLER P ZSINDELY S
6	BONITZ M,EGGHE L FRAME DJ,GOMEZ I MENDEZ A,SNIZEK WE
5	ARUNACHALAM S,ETO H INHABER H,IRVINE J KRETSCHMER H,KUNZ M,LINDSEY D LYON WS,MOED HF NALIMOV VV,PERITZ BC

Table 7.1: The authors ranked according to number of publications

7.5 Who's Who in scientometrics?

Every scientific specialty is carried by human activity. Although possibly not very crucial in understanding the dynamics, knowing "who is who" is not without interest. table 7.1 on page 177 lists all authors that published five or more articles in *Scientometrics* ranked according to their productivity.

The editor of the journal is the most productive author, the adjunct is second of the list. In practice, most authors are part of a small research group. table 7.2 on page 178 shows the most productive institutions ranked in the same way.

A different "who is who" is obtained by looking at the most cited authors. The authors cited more than ten times are shown in table 7.3 on page 179. It should be noted that these are citation counts *within* the data set generated by analyzing the journal *Scientometrics*.

Within this data set, the creator of the *SCI* is the most cited author, followed by Derek Price and the editor of *Scientometrics*.

7.6 Does scientometrics have its own identity?

7.6.1 Method

Scientific specialties can be characterized in terms of patterns in the relationships among the authors of scientific texts. Science is, on the whole, practised in tightly knit communities in which the authors address one another (Crane 1972). To

41	HUNGARIAN ACAD SCI LIB
36	UNIV OREGON
22	LEIDEN UNIV
19	ACAD SCI GDR,ACAD SCI USSR,CSIC
16	UNIV SUSSEX
15	INST SCI INFORMAT,DEPT SCI DYNAM,UNIV INSTELLING ANTWERP
11	COMP HORIZONS INC,DREXEL UNIV,UNIV ILLINOIS
10	BULGARIAN ACAD SCI,GEORGE WASHINGTON UNIV VIRGINIA POLYTECH INST & STATE UNIV
9	ACAD SCI UKSSR,INDIAN INST CHEM BIOL L EOTVOS UNIV,UNIV MONTREAL
8	ECOLE MINES PARIS,OAK RIDGE NATL LAB NATL INST SCI TECHNOL & DEV STUDIES
7	CSIR
6	COLUMBIA UNIV,CORNELLUNIV GEORGIA INST TECHNOL,HEBREW UNIV JERUSALEM NATL SCI FDN,UNIV LEICESTER,UNIV MICHIGAN
5	DSIR,MV LOMONOSOV STATE UNIV UNIV TSUKUBA

Table 7.2: Publishing institutions ranked according to their number of publications

define these patterns in scientometrics, both co-authorship relationships and citations have been analyzed. Co-author relationships can be considered as indicators of co-operation. The meaning of citation relations is less well defined. They can be considered, however, as sociometric data and the resulting network can accordingly be analyzed (Cf. Shrum & Mullins 1988) on its properties.

Authors who published more than one article in *Scientometrics* were extracted from the file with all authors. Since the analysis focused on the *structure* of scientometrics, the "transient" authors (de Solla Price & Gürsey 1976*a*, de Solla Price & Gürsey 1976*b*) were discarded. With these data both a square co-author matrix and a square among-authors citation matrix was produced. The co-author matrix contains by definition symmetric relationships (cell_{ij} is identical to cell_{ji}); the citation matrix contains asymmetric relations. Self-citations were excluded. This procedure results in matrices with mostly empty cells.

The extent to which the authors are connected to one another, i.e. the *cohesiveness* of the network, was analyzed as well as the pattern displayed by each author in relation to all other authors, i.e. the *position* of authors in the network. Authors who do not have a direct relationship with one another can still be quite similar in their pattern of relationships, and consequently hold similar positions in the network. The similarities among authors in both these dimensions of the matrices was also analyzed, i.e. strongly connected authors as well as authors in similar positions were clustered. Direct as well as indirect linkages between the authors were involved in this analysis. A direct relation between author i and j exists whenever cell_{ij} has a value of one. An indirect relationship exists if two authors are related via a third one with whom they have direct relations. For example, in the citation matrix author i will have a direct link with author j if, and only if, i is cited by j. If i is not cited by j, i can still have an indirect link with

240	GARFIELD E
147	DESOLLA PRICE
111	BRAUN T
81	SMALL H
79	SCHUBERT A
65	HAITUN SD
61	MORAVCSIK MI
60	NARIN F
57	CALLON M
55	EGGHE L
50	ARUNACHALAM S
48	BROOKES BC
46	MOED HF
45	COLE S
43	VLACHY J
42	COLE JR
41	FRAME DJ
39	LEYDESDORFF L
35	BONITZ M
30	VINKLER P
29	IRVINE J LINDSEY D MACROBERTS MH
28	BEAVER DD CHUBIN DE
27	BRADFORD SC
26	ALLISON PD MERTON RK
25	COURTIAL JP GRIFFITH BC MARTIN BR
24	CRANE D LAWANI SM
22	CARPENTER MP DOBROV GM
21	ABT HA EDGE D
20	GILBERT GN ZUCKERMAN H
19	LOTKA AJ NEDERHOF AJ
18	GARVEY WD GRUPP H PAO ML YABLONSKI AI
17	GORDON MD MEADOWS AJ NALIMOV VV ORG EC COOP DEV
16	KUNZ M TODOROV R
15	COHEN JE UNESCO
14	COZZENS SE CRONIN B HAGSTROM WO RABKIN YM RIP A
13	IRWIN JO KENDALL MG KRUSKAL JB VELHO L
12	ANDREWS FM BAYER AE BENDAVID J DENNIS W GOFFMAN W INHABER H KRAUSKOPF M KRETSCHMER H QURASHI MM
11	EISEMON TO KUHN TS PELZ DC PRAVDIC N SIMONTON DK

Table 7.3: All authors cited more than 10 times from 1978 until 1993

author j if i is cited by an author who is cited by author j.

Burt's program STRUCTURE was used to analyze these matrices (Burt 1982, Burt & Minor 1983). The basic feature of this program is its ability to analyze matrices with respect to both the relations among the authors (as in graph analysis) and to their position in the network, defined as the pattern of their relations with all other authors in the network. Moreover, the network indices of STRUCTURE are not based on the assumption of a normal distribution of the variables.

For every cell in the matrix, a network index z_{ij} can have one of the following values:

$$z_{ij} = 0 \tag{7.4}$$

if there are no direct or indirect links between i and j;

$$z_{ij} = 1 \tag{7.5}$$

if $i = j$;

$$z_{ij} = 1 - f_{ij}/n_i \tag{7.6}$$

in all other cases.

In this formula, n_i is the total number of authors linked to i, whether directly or indirectly. f_{ij} is the number of all authors linked to i by a number of steps equal to or less than the minimal number of steps needed to link j with i.

The measure z_{ij} varies from 0 to 1.

This network measure provides a basis to analyze differences as distances between the authors in the network. The *relational distance* between i and j, d^{r}_{ij} is computed as:

$$d^{r}_{ij} = 1 - minimum(z_{ij}, z_{ji}) \tag{7.7}$$

This measure has a maximum value of 1 if the i and j are disconnected; it is a measure of "closeness". Second, the *patterns of relations* within the network were measured, in other words, the position of the authors in the network. The formula of the *positional distance* between i and j is in this case:

$$d^{p}_{ij} = \sqrt{\sum_{q}(z_{jq} - z_{iq})^2 + \sum_{q}(z_{qj} - z_{qi})^2} \tag{7.8}$$

In this formula, the summation encompasses all authors in the network. This distance measure approaches zero to the extent that author i and author j have similar relationships with the other authors in the network. The more their patterns of relations differ, the more d^{p}_{ij} increases; it is a measure of "likeness".

Both distance measures were computed for the citation and co-authorship matrices. The authors were subsequently clustered using the subroutine subgroup in STRUCTURE. This clustering can be based on two features: the type of relationship analyzed (relational versus positional) and the strictness of the cluster criterion (strong versus weak). In this way, one can identify strong (relational) cliques, weak (relational) cliques, strong positional or so-called structural equivalence clusters, and weak structural equivalence clusters.

Strong cliques are sets of authors connected by relationships in such a way that all members of the clique are connected to one another, and anyone for whom this holds true is included in the clique. The inclusion criterion is less strong for *weak cliques*, in which all *pairs* within the clique must have relationships with all other pairs, and anyone with a relation to or from a member of the clique is included. *Strong structural equivalence clusters* are sets of authors with completely identical positions in the network (the distance d^p between them is zero). *Weak structural clusters* are sets of authors with a significant similarity in their patterns of relationships (the distance d^p is small).

In order to assess the similarity in weak structural clusters, estimations have to be made with respect to the composition of the weak structural equivalence clusters, i.e. the sets of authors in similar positions. These estimates can subsequently be used for the analysis of covariance matrices for every cluster. Moreover, they can be submitted to factor analysis to test these cluster solutions. In this analysis, only clusters with a loading on one factor of 80 percent or higher were admitted. To prevent taking a sub-optimal solution for an optimal one, an iterative procedure was applied. First, the net was cast as widely as possible to capture all possible members of a cluster. Then all members with a reliability below 0.9 were discarded and the covariance matrices analyzed anew. (Members of strong component clusters have, by definition, a reliability coefficient of 1). This was repeated until a stable solution was reached for every weak structural equivalence cluster of every matrix analyzed.

7.6.2 Results

Of the 669 different authors 73 per cent (488) had published only once. These transient authors are responsible for around 40 per cent of the scientific production in *Scientometrics*. This share remains stable over time. Thus, by focusing on the 181 authors responsible for the remaining 60 per cent of the scientific production in our dataset, we do not introduce distortions when comparing one year with another.

Co-authors

A general phenomenon in science is the growth of the number of co-authored scientific articles, relative to the total scientific production (Cf. Katz et al. 1995). This growth is field specific. In *Scientometrics* 61 per cent of the articles were written by a single author. This share remains stable over time. Apparently scientometrics, unlike the experimental sciences, is still predominantly a solitary affair. This is

Clique 1	Clique 2	Clique 3
BLICKENSTAFF J	BURGER WJM	BAUIN S
BRAUN T	DEBRUIN RE	CALLON M
GLANZEL W	FRANKFORT JG	COURTIAL JP
GOMEZ I	MOED HF	JAGODZINSKISIGOGNEAU M
GRIFFITH BC	NEDERHOF AJ	LAW J
MACZELKA H	PETERS HPF	MICHELET B
MENDEZ A	TIJSSEN RJW	RIP A
MORAVCSIK MJ	VANRAAN AFJ	TURNER WA
MULLINS NC		WHITTAKER J
NAGY JI		
SCHUBERT A		
SNIZEK WE		
TELCS A		
TODOROV R		
WINTERHAGER M		
ZSINDELY S		

Table 7.4: The three biggest co-citation cliques. Apart from these, there are 6 cliques of 3 authors and 16 cliques of 2 authors.

underlined by the large minority of authors who have not co-authored any paper in *Scientometrics*. Overall, this share is 31 per cent, and of the 181 authors who published more than one article in *Scientometrics*, 28 per cent published exclusively single authored papers.

The network of co-authorships is highly fragmented. The number of realized dyadic links among the authors is only a small percentage (5.6 %) of the number of possible dyadic links. The distribution of realized relations is highly skewed, most being either direct relations (i.e. co-authorships) or indirect relations at a distance of one step (i.e. authors who do not co-author with one another but do so with the same third author). With the exception of three subgroups, most co-authors cooperate with no more than one or two colleagues. The clique analysis reveals 16 (strong) cliques of 2 authors, six of 3 and three bigger cliques of respectively 16, 8 and 9 authors (table 7.4 on page 182). Moreover, members of a clique co-author only with members of the same clique. Neither direct nor indirect relationships exist among the cliques.

The analysis of the position of the authors in the network of co-authorship relations reveals 28 strong structural equivalence clusters of 2 authors and 4 clusters of 3 authors, i.e. clusters in which each member has exactly the same position as every other member (table 7.5 on page 183). The analysis shows 10 weak clusters, i.e. clusters of authors with similar positions in the co-author network (table 7.6 on page 184).

Comparison of the composition of the weak structural equivalence clusters with the relational cliques reveals that two clusters are identical: a group of authors from Leiden (Van Raan *et al.*) and a group of authors with various institutional affiliations, probably best characterized as the "co-word analysis group". So, these two groups have distinct identities, with respect both to their relations and to their positions in the network. On the other hand, clique number 1 (Braun *et al.*) turns out to be composed of three groups of authors with distinct positions

Cluster 1	LAW J WHITTAKER J	Cluster 2	MICHELET B JAGODZINSKISIGOGNEAU M
Cluster 3	TURNER WA CALLON M	Cluster 4	CARPENTER MP IRVINE J MARTIN BR
Cluster 5	NOMA E MCALLISTER PR FRAME J	Cluster 6	SWEENEY E GREENLEE E
Cluster 7	PORTER AL STUDER KE	Cluster 8	DIJKWEL PA LEPAIR C
Cluster 9	MANORAMA K GARG KC	Cluster 10	NIEUWENHUYSEN P EGGHE L
Cluster 11	BORDONS M BARRIGON S	Cluster 12	CHATELIN Y ARVANITIS R
Cluster 13	CANO V LIND NC	Cluster 14	LIPATOV YS DENISENKO LV
Cluster 15	MIDORIKAWA N YAMAZAKI S	Cluster 16	ALFENAAR W SPANGENBERG JFA
Cluster 17	KRISHNAIAH VSR NAGPAUL PS	Cluster 18	MILLER RB ZUCKERMAN H
Cluster 19	JIANG GH ZHAO HZ	Cluster 20	BEAVER DD ROSEN R
Cluster 21	BUDD J HURT CD	Cluster 22	KUMARI L SENGUPTA IN
Cluster 23	LONG JS MCGINNIS R	Cluster 24	BLAIVAS A KOCHEN M
Cluster 25	OLUICVUKOVIC V PRAVDIC N	Cluster 26	RABKIN YM INHABER H
Cluster 27	DORE JC FRIGOLETTO L MIQUEL JF	Cluster 28	HASSANALY P DOU H QUONIAM L
Cluster 29	HUSTOPECKY J VLACHY J	Cluster 30	BURGER WJM FRANKFORT JG
Cluster 31	TIJSSEN RJW NEDERHOF AJ	Cluster 32	GOMEZ I MENDEZ A

Table 7.5: Strong structural equivalence clusters in the co-authorship data.

Cluster 1	BAUIN S CALLON M COURTIAL JP JAGODZINSKISIGOGNEAU M LAW J MICHELET B RIP A TURNER WA WHITTAKER J
Cluster 2	CARPENTER MP FRAME J IRVINE J MARTIN BR MCALLISTER PR NARIN F NOMA E
Cluster 3	GARFIELD E GREENLEE E SMALL H SWEENEY E
Cluster 4	BURGER WJM DEBRUIN RE FRANKFORT JG MOED HF NEDERHOF AJ PETERS HPF TIJSSEN RJW VANRAAN AFJ
Cluster 5	BRAUN T GLANZEL W MACZELKA H NAGY JI SCHUBERT A TELCS A ZSINDELY S
Cluster 6	FERNANDEZ MT GOMEZ I MENDEZ A
Cluster 7	BLICKENSTAFF J GRIFFITH BC MORAVCSIK MJ MULLINS NC SNIZEK WE
Cluster 8	DIJKWEL PA LEPAIR C VANHEERINGEN A
Cluster 9	ARUNACHALAM S GARG KC MANORAMA K
Cluster 10	EGGHE L NIEUWENHUYSEN P ROUSSEAU R

Table 7.6: Weak structural equivalence clusters in the co-authorship data.

	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	1	1	0	0	0
6	0	0	0	0	1	1	0	0	0	0
7	0	0	0	0	1	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

CUTOFF AT ZERO

Table 7.7: Block model of relations at subgroup level, defined by positions of co-authorships. A 1 indicates the existence of co-authorship relations between the subgroups, a 0 the absence thereof.

(clusters 5, 6 and 7 of the positional analysis). The positional analysis reveals a different pattern compared with the analysis of only the relations among authors. It creates clusters that would otherwise remain invisible (clusters 2, 3, 8, 9 and 10).

The structure of some clusters seem to be determined by the institutional affiliations of the authors. This is true of the Leiden group and of the authors connected with ISI (cluster 3). In other cases, nationality appears to be the binding factor. This is true of the group in Hungary (cluster 5), the Belgian informetricians (cluster 10) and the Spanish scientometricians (cluster 6). However, cluster 1 can best be characterized by its research program (co-word analysis). Cluster 2 seems to consist of authors from Sussex together with CHI Research Inc. Thus, co-author relations are not only institutionally defined; shared interests and common intellectual goals play a role as well.

Stokes & Hartley (1989) define a specialty as "socially completely cohesive if each and every member co-authors a document with each and every other". A specialty is completely differentiated in the social dimension if no author is a co-author with any other. In these terms, scientometrics is neither completely cohesive nor completely differentiated. It consists of a few big groups of co-authoring authors, many small ones and a large minority (28 percent) of single authors. Among these subgroups no coauthor relations exist. The only exception is the clustering of three different groups of authors in clique 1 (Braun et al.), in which the positional cluster around Braun is the central group. This is shown by the block diagram in table 7.7 on page 185, which gives the relations with structurally similar positions at the level of the clusters (a 1 indicating the existence of co-author relations between clusters).

To sum up these data, scientometrics is a fragmented field of co-authorships.

The authors are highly selective in their co-authorship relations with one another. Co-authorships are defined neither exclusively by social nor only by intellectual factors. Both dimensions seem to shape the pattern of co-authorships. With respect to the number of solitary authors and the large number of isolated small clusters, scientometrics exhibits the pattern of a normal social science.

Citations

Of the 779 articles published in *Scientometrics*, 411 were subsequently cited one or more times in *Scientometrics*. The share of references to *Scientometrics* (as a percentage of all references) has stabilized around an average of 19.4 percent since 1987. Of the 181 authors in the core set, 130 authors cite one another. This means that 51 (or 28.2 percent) of the authors publishing more than one article in *Scientometrics* from 1978 till 1993 are neither citing nor cited within this group of authors.

The core set of authors in *Scientometrics* is found to be highly cohesive in terms of their mutual citation relations. All these authors are members of one single weak clique. Moreover, a majority of these authors (88) also belongs to one strong clique (table 7.8 on page 187). The picture is different if we exclude all indirect relations from the analysis. This "fine structure" of the citation matrix is shown in table 7.9 on page 188 and table 7.10 on page 189, where 13 strong cliques and 6 weak cliques are revealed. Most strong cliques seem to coincide with shared institutional affiliations. The exception is clique 9, which indicates the existence of a debate among the members of this clique. The cohesiveness of the citation network of *Scientometrics* is, given the strict condition that only direct relationships are analyzed, underscored by the fact that no fewer than 31 authors cluster together (in weak clique 6 in table 7.10 on page 189). This cohesiveness is also apparent from the distribution of relationships actually realized as a percentage of the possible dyadic relationships in the citation network: only 4 percent of all possible citing relations are realized, although most authors are connected indirectly to each other at distances of 2, 3 and 4 steps.

If we analyze the *pattern* of citation relations, the mutual citation matrix of *Scientometrics* seems to consist of 8 different sets of authors. If one wishes to know which authors are similar in their patterns of citing and being cited, one must exclude all indirect relations within the matrix from the analysis. The result is shown in table 7.11 on page 190. The majority of authors cluster together on the criterion of their position in the network (cluster 3). It can be concluded that most scientometricians are similar in terms of their direct citation relationships within this group. This is true regardless of the extent to which an author is connected to other authors. Most positional clusters do have citation relations with one another, which indicates an integration of the citation network at the subgroup level.

In summary, within the core set of 181 authors in *Scientometrics*, 130 authors cite one another regularly. The majority have strong mutual relationships, being members of one clique. Moreover, the pattern of citation relations of most authors is very similar. This is even true of the relatively isolated authors. Thus,

ADAMSON I	GORDON MD	NOMA E
ARUNACHALAM S	GRANOVSKY YV	OLUICVUKOVIC V
BALDAUF RB	GREENLEE E	OROMANER M
BAUIN S	HAITUN SD	PERITZ BC
BLICKENSTAFF J	HARGENS LL	PRAVDIC N
BONITZ M	HUSTOPECKY J	RABKIN YM
BRAUN T	INHABER H	RIP A
BURGER WJM	IRVINE J	ROUSSEAU R
CALLON M	KRETSCHMER H	SCHUBERT A
CARPENTER MP	KUMARI L	SEN SK
CHUBIN DE	KUNZ M	SENGUPTA IN
COHEN JE	LANCASTER FW	SMALL H
COURTIAL JP	LANGE L	SMART JC
COZZENS SE	LEPAIR C	SNIZEK WE
DAVIS CH	LEYDESDORFF L	SWALES J
DEBRUIN RE	LINDSEY D	SWEENEY E
DIAMOND AM	LUUKKONEN T	TAGUE J
DIJKWEL PA	MANORAMA K	TIJSSEN RJW
DOREIAN P	MARTIN BR	TODOROV R
EGGHE L	MCALLISTER PR	TURNER WA
ETO H	MENDEZ A	VANRAAN AFJ
FERNANDEZ MT	MILLER RB	VELHO L
FRAME J	MOED HF	VINKLER P
FRANKFORT JG	MORAVCSIK MJ	VLACHY J
GARC KC	NAGY JI	WINTERHAGER M
GARFIELD E	NARIN F	ZSINDELY S
GLANZEL W	NEDERHOF AJ	ZUCKERMAN H
GOMEZ I	NIEUWENHUYSEN P	

Table 7.8: The strong component clique in the citation data.

Cluster 1	HARGENS LL LINDSEY D
Cluster 2	BONITZ M KUMARI L SENGUPTA IN
Cluster 3	DEBRUIN RE NEDERHOF AJ TIJSSEN RJW
Cluster 4	GREENLEE E SMALL H SWEENEY E
Cluster 5	BAUIN S CARPENTER MP IRVINE J MARTIN BR
Cluster 6	TODOROV R WINTERHAGER M
Cluster 7	FRAME J MACALLISTER PR NARIN F NOMA E
Cluster 8	BURGER WJM FRANKFORT JG MOED HF VANRAAN AFJ
Cluster 9	CALLON M COURTIAL JP LEYDESDORFF L TURNER WA
Cluster 10	HAITUN SD KUNZ M
Cluster 11	ARUNACHALAM S GLANZEL W
Cluster 12	MORAVCSIK MJ VELHO L
Cluster 13	BRAUN T SCHUBERT A ZSINDELY S

Table 7.9: Citation matrix: cliques using only direct citation relations

Cluster 1	HARGENS LL LINDSEY D
Cluster 2	GREENLEE E SMALL H SWEENEY E
Cluster 3	CALLON M COURTIAL JP LEYDESDORFF L RIP A TURNER WA
Cluster 4	HAITUN SD KUNZ M
Cluster 5	MORAVCSIK MJ VELHO L
Cluster 6	ARUNACHALAM S BAUIN S BONITZ M BRAUN T BURGER WJM CARPENTER MP DEBRUIN RE EGGHE L FRAME J FRANKFORT JG GARG KC GLANZEL W IRVINE J KUMARI L MACALLISTER PR MANORAMA K MARTIN BR MOED HF NARIN F NEDERHOF AJ NIEUWENHUYSEN P NOMA E ROUSSEAU R SCHUBERT A SENGUPTA IN SWALES J TIJSSEN RJW TODOROV R VANRAAN AFJ WINTERHAGER M ZSINDELY S

Table 7.10: Weak component cliques using only direct citation relations.

Cluste 1	BLAIVAS A, DOREIAN P, KOCHEN M KRETSCHMER H, LANGE L, SWALES Jr
Cluster 2	BONITZ M, EGGHE L, NIEUWENHUYSEN P ROUSSEAU R, TAGUE J
Cluster 3	ADAMSON I, ALFENAAR W, ARUNACHALAM S, ARVANITIS R BALDAUF RB, BARRIGON S, BAUIN S, BEAVER DD BLICKENSTAFF J, BORDONS M, BRAUN T, BRUNK GG BURGER WJM, CALLON M, CANO V, CARPENTER MP CHATELIN Y,CHU H, CHUBIN DE, COHEN JE COURTIAL JP, COZZENS SE, DAVIS CH, DEARENAS JL DEBRUIN RE, DIAMOND AM, DIJKWEL PA, DORE JC DOU H, EHIKHAMENOR FA, ETO H, FERNANDEZ MT FRAME J, FRANKFORT JG, FRIGOLETTO L, GARFIELD E GARG KC, GLANZEL W, GOMEZ I, GORDON MD GRANOVSKY YV, GREENLEE E, GRUPP H, GUAY Y GUPTA DK, HAITUN SD, HARGENS LL, HASSANALY P HURT CD, HUSTOPECKY J, INHABER H, IRVINE J AGODZINSKISIGOGNEAU M, JASCHEK C, KORENNOI A KRAUSKOPF M, KRYZHANOVSKY LN, KUNZ M, LANCASTER FW LAWANI SM, LEMOINE W, LEPAIR C, LIND NC, LINDSEY D MACALLISTER PR, MACZELKA H, MANORAMA K, MARTIN BR MCCAIN KW, MENDEZ A, MICHELET B, MILLER RB MIQUEL JF, MOED HF, MORAVCSIK MJ, NAGY JI NALIMOV VV, NARIN F, NEDERHOF AJ, NOMA E NORDSTROM LO, OROMANER M, PAO ML, PERITZ BC PLOMP R, POURIS A, QUONIAM L, QURASHI MM RABKIN YM, RICHARDS JM, RIP A, ROCHE M ROSEN R, RUSHTON JP, RUSSELL JM, SEN SK SMALL H, SMART JC, SNIZEK WE, SPANGENBERG JFA SWEENEY E, TIJSSEN RJW, TODOROV R, TROFIMENKO AP TURNER WA, VANHEERINGEN A, VELHO L, VINKLER P VLACHY J, WINTERHAGER M, YABLONSKY AI, YUTHAVONG Y ZSINDELY S, ZUCKERMAN H

Table 7.11: Authors with similar positions: citation matrix analyzed only on structural equivalence in the direct citation relations.

1178	SCIENTOMETRICS
331	J AM SOC INFORM SCI
248	SCIENCE
201	RES POLICY
190	SOCIAL STUDIES SCI
157	SOC STUD SCI
131	NATURE
122	AM PSYCHOL
114	AM SOCIOL REV
102	J DOCUMENTATION
101	CZECH J PHYS B
85	CURRENT CONTENTS
84	LITTLE SCI BIG SCI
75	J DOC
66	J INFORM SCI
64	AM J SOCIOL
47	SCI STUDIES
41	ESSAYS INFORMATION S SOCIAL STRATIFICATIO
39	MINERVA
35	HDB QUANTITATIVE STU
34	SCI AM
33	AM DOCUMENTATION
32	AM SOCIOL PSYCHOMETRIKA
30	HIST SCI J WASHINGTON ACADEMY
29	SOCIAL SCI INFORMATI
28	EVALUATIVE BIBLIOMET INTERCIENCIA J INFORMATION SCI J POLITICAL EC
27	CITATION INDEXING IT INFORM PROCESS MANAG INFORMATION PROCESSI MAPPING DYNAMICS SCI

Table 7.12: The most cited journals

these results are indicative of a strong integration. In this respect, the field of scientometrics seems indeed to have developed a social identity of its own.

7.7 What is scientometrics' position?

The position of *Scientometrics* can be analyzed by looking at the journals that are cited by *Scientometrics*. This is actually a quantitative "self-portrait", displayed in table 7.12 on page 191.

Apparently, scientometrics as represented by its core journal, is most intimately related to library and information sciences, science studies and science policy studies.

7.8 Has scientometrics developed a specific language?

7.8.1 Method

The question of whether scientometrics has developed a specific language of its own, may be rephrased as follows: has scientometrics caused its own specific semantics to emerge? This is in no way self-evident. After all, it might also be the case that the articles in *Scientometrics* use various different terminologies employed by the disciplines surrounding scientometrics. If this were the case, one would expect distinct sets of words to be used with few or no connections between them. One language for all scientometricians would on the other hand be indicated by strong connections between the words used.

Given the functions of titles of articles, the words in these titles can be considered as indicators of the cognitive message of the publication (Rip & Courtial 1984, Leydesdorff 1989). The co-occurrence of words in titles can be considered to be an indication of the existence or non-existence of relations between these words (Callon et al. 1983). Consequently, co-occurrence data can be analyzed as sociometric choice data and subjected to network analysis.

All titles and title words were extracted From the file for the period as a whole and also for every three year period (1992-1990, 1990-1988, 1988-1986, 1986-1984, 1984-1982, 1982-1980, 1980-1978). Word versus title matrices were computed, in which $cell_{ij}$ = 1 if word j occurs at least once in title i. From this matrix a square word-word matrix was obtained where $cell_{ij}$=1 if words i and j occur in the same title at least once. This matrix was subsequently subjected to the network analysis referred to above, producing strong cliques, weak cliques, strong structural equivalence clusters and weak structural equivalence clusters. In separate runs, the direct relationships between the words in these matrices were additionally analyzed by filtering out all indirect links.

7.8.2 Results

The most striking feature of the network of title words of articles published in *Scientometrics* is its cohesiveness. All words cluster together in a single strong component clique (table 7.13 on page 193). If only direct relations are included, all words cluster together in a single weak component clique. This means that all words are either used together in a title or share a common co-word. This strong cohesivess is underlined by table 7.14 on page 194, which gives the realized word-word relations as a percentage of all possible dyadic relations for the matrix as a whole and for every three year period. Almost all the co-word possibilities are realized in two steps, meaning that the words are at the most one step apart. So, the titles exhibit many overlaps.

At the same time, the language of scientometrics appears not to be strictly codified. The words display a highly individual pattern in their relations to one another. In contrast to the authors, most words cannot be clustered reliably at all. Even if the threshold of reliablity coefficients is lowered to 0.85 (instead of

ACTIVITIES	EVALUATION	PROBLEM
ACTIVITY	FIELD	PRODUCTIVITY
ANALYSIS	GROUP	PROGRAM
APPLICATION	GROWTH	PUBLICATION
ARTICLE	IMPACT	QUANTITATIVE
ASSESSMENT	INDEX	RESEARCH
AUTHORSHIP	INDICATOR	REVIEW
BASIC	INFORMATION	SCIENCE
BIBLIOMETRIC	INTERNATIONAL	SCIENTIFIC
BRADFORD	JOURNAL	SCIENTIST
CASE	LAW	SCIENTOMETRIC
CHEMISTRY	LIFE	SIZE
CITATION	LITERATURE	SOCIAL
CO	MEASURE	STATE
COLLABORATION	MEASUREMENT	STRUCTURE
COMMENT	METHOD	STUDIES
COMMUNICATION	MODEL	STUDY
COMPARATIVE	MULTIPLE	SYSTEM
COMPARISON	NATIONAL	TECHNOLOGICAL
COUNTRIES	NETWORK	TECHNOLOGY
DATA	OUTPUT	THEORY
DEVELOPMENT	PAPER	UNITED
DISCIPLINE	PATTERN	UNIVERSITY
DISTRIBUTION	PERFORMANCE	USE
ECONOMIC	PHYSICS	USING
EFFECT	POLICY	WORLD
EUROPEAN	PRICE	

Table 7.13: The overall network of title words

0.9 as used in the analysis of the author-author matrices), only two small clusters of words in similar positions show up (table 7.15 on page 194). Apparently, the network of title words does not have a clear structure. This may be explained by postulating that there is no clear subdivision of the set of words used in titles in *Scientometrics*. Nor do distinctions between subject, methodology and theory-related words show up in the analysis.

Thus, the language of scientometrics is both strongly unified and weakly codified. This strong cohesiveness has been a stable characteristic of the titles in *Scientometrics*, from the very start of the journal's existence. This seems to suggest that a distinct discourse was already in existence before the journal was actually founded. In any case, it constitutes a textual identity of scientometrics as a field, one probably different from the various mother disciplines. Thus a process of de-differentation seems to have occurred not only in the patterns of citing (and being cited) but also on a cognitive level.

7.9 Conclusion

In more than one respect, *Scientometrics* displays the characteristics of a social science journal. Its Price Index is in the domain of the library and information

PATH DISTANCE	FREQUENCY	PERCENTAGE
1	2852	45.13
2	3468	54.87
3	0	.00
NO CONNECTION	0	.00

Table 7.14: Realized word-word relations as percentage of possible dyadic relations 1978-1992

Cluster 1	Cluster 2
NATIONAL INTERNATIONAL	ANALYSIS INDICATOR CITATION SCIENTIFIC RESEARCH SCIENCE

Table 7.15: Words with similar positions in the epistemic network.

sciences. The majority of its published items are written by a single author. Moreover, the network of co-authorships is highly fragmented: most authors cooperate with no more than one or two colleagues. This fragmentation does not preclude the citation networks and the title networks from displaying a remarkably high cohesion.

This cohesion may be interpreted as a bibliometric indication that the field, as represented by its core journal, has indeed developed an identity of its own. The strong fragmentation in the co-authorship network may be seen as resulting from fierce competition between the research groups for contract research related to the development and application of science and technology indicators.

Back to Contents Page

Back to Contents Page

Chapter 8

Representing science

8.1 Introduction

A recurring theme in the use of science and technology indicators, as well as in the construction of new ones, has been the interpretation of these indicators. Given the dependence on citation data of the majority of interesting science and technology indicators, a general citation theory should make the meaning of science and technology indicators more transparent. Hence the continuing call for a citation theory in scientometrics and, sometimes, in science policy. So far, such a theory has not yet been developed by the experts in the field. This study suggests an explanation for this. In the final chapter, the solution I propose to the problem of citation theory will be set out in more detail. I will furthermore try to use this study's results to better understand the ways in which the sciences are represented, both in the process of knowledge production itself and in science policy.

8.2 Summary of the results so far

This study defines and analyzes the citation culture by theoretically distinguishing the citation from the reference. Chapter 1 argues that it makes sense to analyze the reference and the citation on their capacity as signs. Since the citation and the reference have different referents and are actually each other's mirror image, it does not seem very wise to blur the analytical distinction between them. This distinction has moreover the advantage that the quest for a citation theory in scientometrics and the sociology of science splits into two different, analytically independent research problems: the patterns in the citing behaviour of scientists, social scientists and scholars in the humanities on the one hand, and the theoretical foundation of citation analysis on the other. As a consequence, the locus of the birth of the citation shifts from the scientist's desk to the indexer's office. The question "how are these citations produced?" becomes more relevant than it would have seemed if this study had stuck to the conventional wisdom that the distinction between reference and citation is merely technically interesting.

Chapter 2 and chapter 3 study the production process of the citation by describing, on the basis of documents and interviews, the creation and building of the first *SCI*. The distinction made in chapter 1 proves fruitful: Eugene Garfield and Joshua Lederberg turn out to have accomplished the translation of Shepard's legally bound citation into a new sign of science. Chapter 2 and chapter 3 also clarify the otherwise puzzling rejection of the *SCI* by many scientists. This resistance was inherently present from the very beginning and did not need any misuse before emerging. This is difficult to explain if one assumes, as most authors on the topic have done, that citation analysis is a more or less "natural" or plausible extension or use of the citation norms in science and the social sciences.

Chapter 4 follows up on this historical analysis by sketching some consequences of *SCI*'s publication. It describes the way the *SCI* was received by sociologists of science on the one hand, and by the science of science movement on the other. It also relates, on the basis of original Russian and Ukranian publications and interviews, how the *SCI* contributed to the early emergence of two different types of scientometrics in the Soviet Union, one oriented towards natural scientists and the other to science policy institutions. The chapter argues that the response to the *SCI* differed. The Mertonian sociologists of science — who together with Derek de Solla Price were the first to systematically study the potential of the *SCI* as a sociological research instrument in their graduate seminar at Columbia University — incorporated citation data in their studies where they seemed to fit in with other data and with their theory. The science of science movement saw the *SCI* as its dream coming true: science would finally be enabled to analyze itself. This was moreover to provide the foundation for a rational science policy. Thus, the *SCI* was used in two different settings, Derek de Solla Price being one of the few who played a role in both contexts. The "capture" of the science of science tradition by the upcoming citation culture (embodied in the *SCI*) resulted in the social science specialty of scientometrics. This also explains why scientometrics has had a distinct position in the whole of science studies from the very beginning, in a perhaps somewhat subtler way than the usual distinction between those people who hate numbers and those who love them.

The *SCI* did, however, not only give rise to the specialty of scientometrics, it also laid the foundation of a whole set of new indicators of science and technology. Chapter 5 tries to spell out precisely how the citation is used as a building block of the new indicator building. The chapter also tries to capitalize on the distinction made in chapter 1 by looking at how the two different dimensions of "citing" and "being cited" are interacting in the shaping of novel indicators. In this way, chapter 5 shows the analytical fruitfulness of the point of departure of this study in the deconstruction of indicators. By analyzing in detail the most strategic types of indicator, i.e. the ones that are dominant in citation analysis, the implicit claim of the chapter is that this deconstruction can be done in essentially the same way on any scientometric or bibliometric indicator. Not so surprisingly, the chapter also unveils the "heterogenous engineering" character of the whole enterprise. Whereas indicators tend to be judged on their consistency, the creation of the prevailing scientometric indicators turns out to be a far from consistent process. Although this certainly does not invalidate their use, it does elucidate their character as semiotic and socio-cognitive products of creative scientometricians.

Tinkering with indicators is also a characteristic of science policy, as chapter 6 has shown. In this domain, the demand for measuring scientific performance acted as a catalyst. Science funding bodies, sometimes confronted with the conflict between a lack of budgets and rising numbers of good research proposals, wished to know how effective their funding policies were. This was the context in which Lederberg, sitting in at an NIH meeting in 1957, recalled the article by Gene Garfield in *Science* some years before. Essentially the same problem created the first Dutch citation analyses in the physics funding body FOM. In its turn, this stimulated the Science Advisory Council to develop an explicit indicator policy in its drive to open up the bastion of science in the Netherlands. Ultimately, this led to one of the largest scientometric centres in the world at Leiden University.

This pattern, in chapter 6 retold with respect to Dutch science policy, is not unique to the Netherlands. For example, a closely related question, "How well do big laboratories perform?", was the research problem which led to the highly controversial performance indicators developed at the Science Policy Research Unit Sussex University at the end of the 1970s[1]. The problem how well state-funded chemistry laboratories were performing led to co-word analysis at the Paris Ecole des Mines in 1976[2] Apparently, different political contexts, different periods, and different actors gave rise to very similar patterns: the creation of specific performance indicators and related scientometric methodologies. Without a demand for science and technology indicators, scientometricians would not have been urged to mass produce indicators. Science and technology indicators would probably still have existed, but would not have acquired the independent status they enjoy nowadays. In short, whereas chapter 4 and chapter 5 deal with the push emanating from the new symbols embodied in the *SCI*, in chapter 6 the building up of a market for scientometric indicators has been sketched.

Chapter 7 has tried to take a quantitative look at the core journal of scientometrics, *Scientometrics*. The empirical question in chapter 7 is a simple one: how has the specialty developed if we take the first 25 years of its core journal as indicative? The analysis shows first of all that notwithstanding its being born out of the fusion of the scientistically inclined science of science and the *SCI*, scientometrics displays the characteristics of a social science, not of a natural science. In this sense, Price's dream of a more objective type of sociology, resembling the natural sciences, does not seem to have come true[3]. On the other hand, scientometrics' position in the domain of information science positions the field in the area of the relatively "hard" social sciences, as perceived by their practitioners.

Chapter 7 also shows that scientometrics, as represented by its core journal, does seem to have specific properties as a specialty. Whereas the patterns of both the co-author relationships and the citation network hint at scientometrics as a specific domain of scientific activity, the analysis of the title words does not reveal specific patterns. The words simply clump together in one big cluster. Apparently, scientometricians all speak the same language. Yet, research groups do not collaborate very strongly. This can be explained by the features discussed in

[1] Ben Martin, Interview, Brighton.

[2] Michel Callon, Interview, Paris.

[3] This conclusion contradicts Schubert & Maczelka (1993).

chapter 6 and chapter 5. Access to clean data is a strategic asset and is always jealously guarded, not least because of the abundance of contract research. Cordial relationships do indeed exist between different research groups, but data and methods of using them are not freely available.

8.3 A hybrid specialty

Since the seventies, several models of specialty development have been proposed in science studies. Most of these concentrated on successful and homogeneous academic specialties. For this reason the received specialty models do not seem well suited to deal with the hybrid entity this study deals with. In the constructivist turn within science studies, they have moreover been criticized for their naive way of demarcating specialties. It may be better to combine insights from the older specialty models with those of the later constructivist studies which zoom in on scientists' heterogeneous engineering. This leads to a heterogeneous model in which the development can be represented as being driven by both a pull and a push mechanism.

The push force would be provided for by the unfolding of the citation culture after the birth of the *SCI*. Since 1964, the citation sign has been mass produced on a regular basis (the *SCI* is published four times a year). Because of its specific semiotic properties and its relationship with various citing cultures in science, this created the possibility of a host of interrelated indicators (chapter 5). This did not materialize of its own accord, though. Signs generally do not move by themselves. The citations needed symbolic actors (people) for whom these symbolic possibilities were also social and economic opportunities to perform research and thereby create or sustain careers. The pull force would be generated by science policy, the central theme of chapter 6. Since science policy has passed the stage at which it merely distributed money for research, it has encountered an annoying paradox. To judge the state of affairs in research, it depends on the very expertise it is supposed to steer and shape. The scientometric indicators seem to provide a solution: they enable expert judgement which is not dependent on the scientists involved. This promise of solving science policy's paradox created a novel market for quantitative science and technology indicators, a process studied in chapter 6.

Thus, the push and a pull mechanisms would interact with one another in the building up of science and technology indicators and in the shaping of the specialty of scientometrics.

8.4 Indicators as translators

8.4.1 Science as an information cycle

The implications of the emergence of both scientometrics and a set of interconnected science and technology indicators can be drawn out by taking a closer

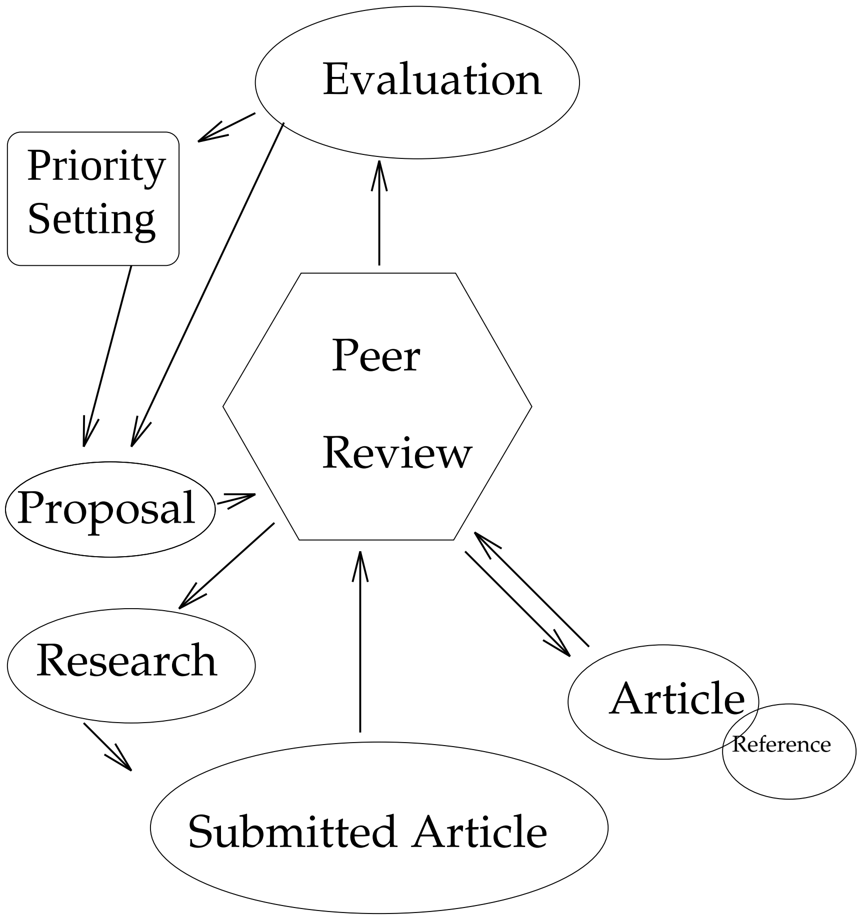

Figure 8.1: The peer review cycle

look at the process of knowledge production. The production of scientific knowledge can be pictured as a cyclical process in which certain inputs, like money and labour force, are translated into certain outputs, like scientific articles and knowledge claims. The various steps in this process have been analyzed in science studies, both at the micro level (Knorr-Cetina 1981, Latour & Woolgar 1986), at the meso level (Lemaine, MacLeod, Mulkay & Weingart 1976), and at the macro level (Jasanoff 1990, Latour 1984). An important feature is the constant evaluation of scientific knowledge. This assessment is used in the setting of new targets and the writing of new research proposals. Figure 8.1 on page 199 gives a schematic overview of this cycle. As can be seen, peer review in its various forms is central to this cycle. This is the reason for naming it "the peer review cycle".

The peer review cycle itself is the product of a rather complex and convoluted history. It should not be seen as science in its purest form, since it has itself been heavily influenced by science policy considerations in the past (Cozzens, Healy, Rip & Ziman 1990). Nevertheless, it is widely felt to be one of the most distinctive features of science. If the description of this cycle is, rather arbitrairily, started with the writing of a research proposal, the second step is the proposal's evaluation by those peers involved (using both scientific and extra-scientific criteria). The research is then carried out, possibly leading to submission of an article. A second form of peer judgement, organized by the editor of a journal, produces, if positive, a publication. A third form of peer review is the assessment, at regular intervals, of research groups, university institutions and even the national contributions to specialties as a whole. These evaluations are again based on expert judgement. The results of these diverse evaluations contribute to the process of priority setting at various levels, leading to new proposals for research, instititutional transformations or priority programs. The whole of peer review procedures is part of the reputational control system, as analyzed by Whitley (1984).

This is, of course, a rather general representation. The different stages in this cycle are not always executed and they may take various forms in different countries. It should moreover not be read as the claim that science proceeds in a single turn of the cycle. It is an abstract representation of a multitude of interlocking procedures. The scheme does represent, however, important properties of knowledge production. In representing science this way, "information" is taken as the entity that flows and as the substance that is translated in various forms during this cycle. It can easily be seen that this peer review cycle is in no way autonomous. For example, political priorities influence the step from evaluation to priority setting and to a growing extent also the peer judgement of research proposals by research councils. Monetary and economic arguments influence the overall science budget. And the peer judgement as well as the formulated scientific problems are contingent on culture at large. Nevertheless, the scheme shows the central position of scientific expertise. Whereas a large number of social and cultural factors influence the evaluation of science, the resulting quality judgement has to be justified in terms of the digestion of the cognitive products of science, as laid down in the scientific literature.

The dominant role of relevant domain-bound expertise is changed by the bibliometric indicators. Since their advent, a scientific publication can be measured by citation analysis or positioned by co-word analysis. Hence, the expert in the field is no longer the sole source of evaluation expertise. In terms of the information cycle representation, the bibliometric indicators appear in the form of an new, added cycle. This cycle processes information about the primary information cycle, i.e. the peer review cycle. Apparently, this secondary cycle produces and transforms information about information, meta-information. Hence, the secondary cycle is also an information cycle.

Contrary to the first cycle, the secondary cycle does have a distinct beginning. This is the consequence of its contingency on the primary information cycle. Its first step is the semiotic inversion of the reference into the citation (chapter 1). The citation indicators are the main pillars upon which maps of science are built. A slightly different translation process takes place in co-word analysis. In this case, the article is translated into a set of keywords (indexer determined keywords, selected title words or a selection from the abstract or full text). These keywords are subsequently used to construct co-word indicators. Again, maps of science can be created with the help of these indicators. It is also possible to combine co-word indicators and co-citation indicators to construct maps of science.

If citation analysis were simply a numerical mirror of peer review judgements, nothing much new would be happening. This is, however, not the case (chapter

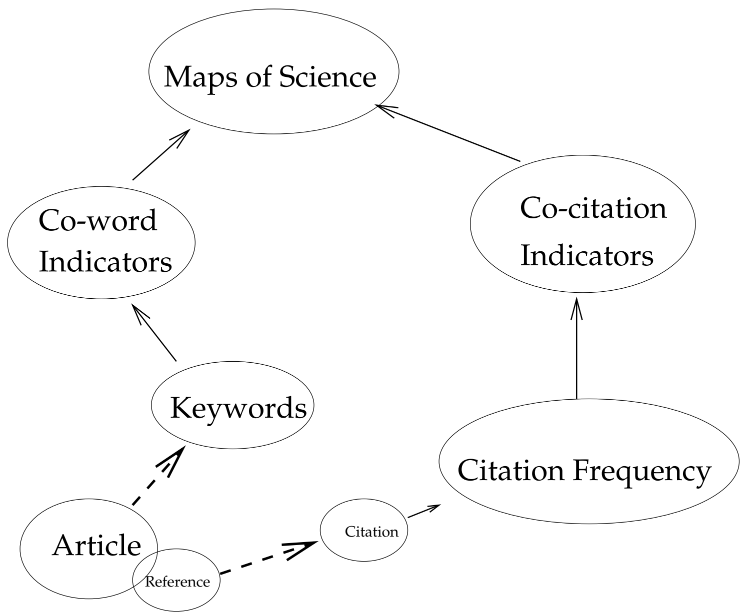

Figure 8.2: The citation cycle

5). The translation of references into citations creates additional degrees of freedom in handling citations. This is enhanced by the construction of more elaborate indicators and maps, all of which implicate numerous more or less arbitrary decisions. This does not mean that anything goes. It does mean, however, that functionally equivalent indicators can be made in several ways, as discussed in chapter 5.

Since the citation plays a crucial role in the second cycle, this cycle may be called "the citation cycle". This needs some explanation. Derek de Solla Price was the first to formulate this concept (Price 1979). The cycle I am proposing is rather different from Price's. The father of scientometrics wished to

> exhibit an interlocking metabolic complex of bibliometric (and scientometric) parameters in a comprehensive and integrated structure after the manner of the Nitrogen Cycle. (Price 1979, 621)

Price's citation cycle follows the construction of the *Science Citation Index* itself and tries to quantify the relationships among such items as source authors, cited authors, source publications and citing and cited articles. Notwithstanding Price's witty and clever text, it is not exactly clear what precisely cycles in his citation cycle. It looks more like a tourist streetcar route illustrating various aspects of the structure of science than like a metabolic cycle. Sentences like "For the next stage in the tour we enter the domain of citations" (Price 1979, 625), seem to indicate that this was exactly what Price intended. The cycle I am proposing is, on the contrary, a dynamic cycle in which information about the production of knowledge is processed and transformed. Its main results are representations of science: in the form of information on the performance of researchers, research institutions or other actors in terms of certain indicators; in the form of maps of science; and in the form of ratings of, for example, journals in terms of impact factors.

8.4.2 Interactions between the cycles

The policy discussions as well as the validation of indicators by the experts in the fields to be assessed are based on translations of concepts in the domain of one cycle into concepts relating to the other. The following types of interaction between the cycles can be hypothesized to occur[4]:

1. The indicators may influence the evaluation of science as such. This is for example the case whenever the citation frequency is used as a measure of scientific performance.

2. The indicators may in a more indirect way redefine the notion of quality in the realm of peer review judgements. This is, for example, the case whenever publication in a source journal of the *SCI* becomes an independent criterion in the assessment.

3. The maps of science (and other complicated science representations) may alter the evaluation process. These maps may for example influence the mental map of scientific experts judging a certain sub-specialty.

4. Scientific experts may be involved in the choice and validation of bibliometric indicators, and co-citation and co-word maps. As has been shown, this has become something of a routine procedure. Since the maps involved are highly sensitive to the thresholds used at different stages in the computational process, the measures used and their resulting maps may be fine-tuned to produce images that make sense to experts in the field. These validated maps may subsequently be used in follow-up analyses.

5. Scientists may also be involved in the validation of ranking lists. It is imaginable, for example, that experts can pinpoint anomalous ranking phenomena, if only because they know the people on the list (contrary to the scientometrician).

6. Scientific experts may also be directly involved in the construction of quantitative indicators. They may, for example, be used as a source of expert knowledge on specific features of the scientific literature involved.

All forms of interplay between the two information cycles have in common that they entail a translation of one type of science representation into another. In figure 8.3 on page 203, the upper right half represents the domain of the citation cycle with its formal representation of the scientific literature. The lower left half represents the domain of the peer review procedures with its stressing of the cognitive dimension of science. Since meta-information cannot in itself be

[4]Empirical research to what extent these interactions actually have occurred is outside of the scope of this study.

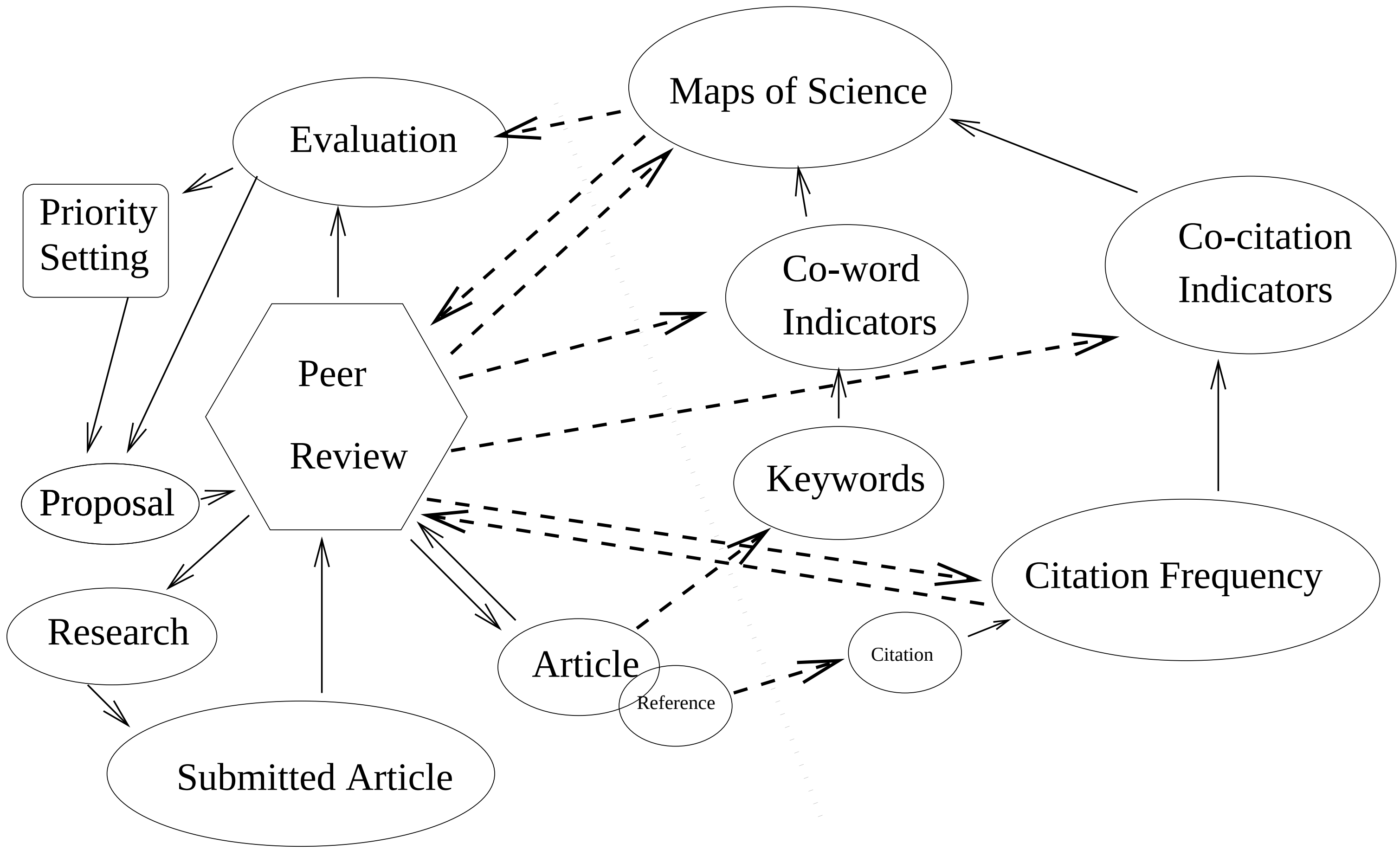

Figure 8.3: The cycle interactions

distinguished from information, the two cycles may interact easily. Science policy tends to promote this interplay. This means that the two cycles will tend to change the very foundations on which they are built. In other words, neither can easily be found by empirical means in their pure form. Consequently, influence of one on the other will tend to go unnoticed.

These interactions may influence the scientific system at all levels, from the individual scientist to the realm of science policy. First of all, the evaluation of science may now get two different kinds of input: one representing the conclusion of the field-specific expert and one representing scientometric expertise. The confrontation between these two forms of expertise has been a main feature of policy debates on measuring science. Because of this, the field-specific scientist no longer has a monopoly position in evaluating science. The two different forms of evaluation do not, by the way, have to be distinguishable in any clear-cut fashion. The various interactions mentioned above, will on the contrary promote the blending of the two perspectives. Nevertheless, they represent analytically different science representations. These differences create both the space and the need for negotations and mutual validations between the expert opinion of the scientists and the scientometric expertise. Incidentally, the effort these translation processes cost is itself an indication that citation analysis is far from identical to peer review.

8.4.3 Credibility cycles

The citation cycle may also transform the way scientists earn recognition. These reward processes can also be represented with the help of cycles (Latour & Woolgar 1986, Knorr-Cetina 1981, Cozzens et al. 1990). Figure 8.4 on page 205 shows the credibility cycle as discussed in science studies (Latour & Woolgar 1986, Rip 1996). This sketch presumes an identity between the reference and the citation. Since this is no longer tenable, I propose an adapted cycle, shown in figure 8.5 on page 206, which takes the consequences of the citation cycle into account.

Since measuring performance indicators is based on fundamentally different expertise from judging the intellectual novelty of a paper, the credibility cycle bifurcates. A new loop is added, making the credibility cycle more complex. The appearance of scientometrics in these credibility cycles may indeed be the main cause of the need of an ethical consciousness in scientometrics: indicators have the potential to end as well as make careers. Of course, the extent to which scientific credibility is made dependent on quantitative indicators or on qualitative judgements may vary widely over different institutions and scientific cultures.

8.4.4 Implications of the citation cycle

The interactions mentioned above are all feasible, although empirical research of these interactions should clarify the extent to which they are actually changing the creation and use of scientific knowledge. In any case, scientometricians are often eager to include scientific experts in their validation work. This has been

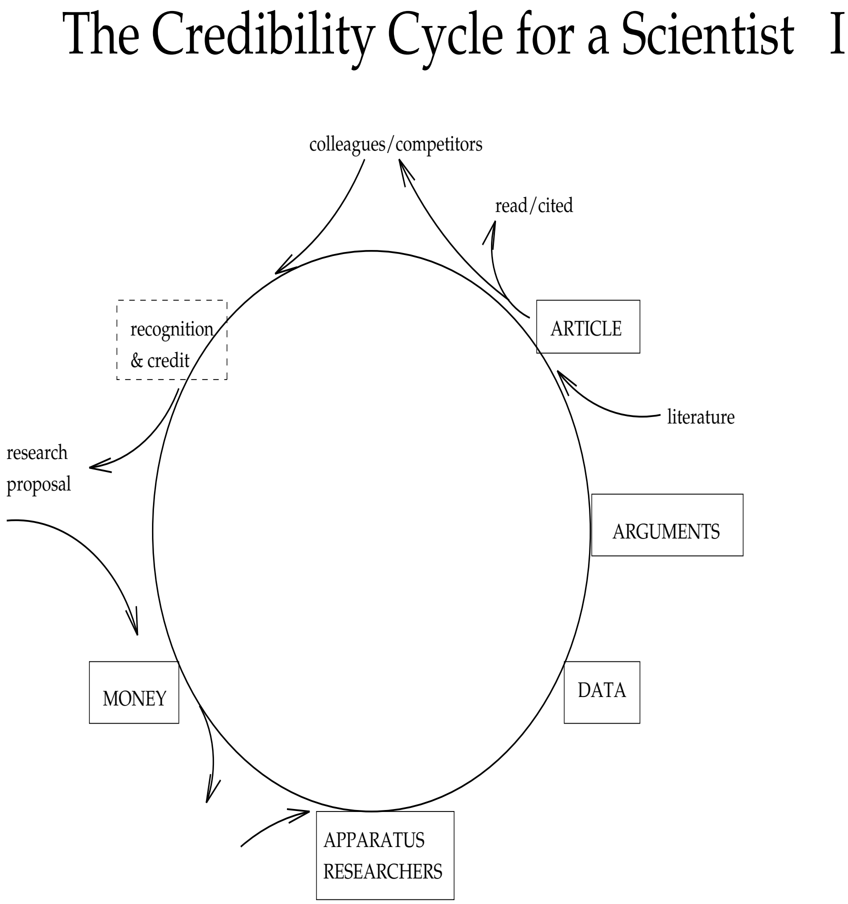

Figure 8.4: The classical credibility cycle

the case from the very beginning of scientometric mapping (Narin 1976). The same is true of the inclusion of scientists in the interpretation of maps of science and even in the construction of fine-grained indicators[5]. The citation cycle also seems to influence the peer review cycle, although the record of evaluating specialties and disciplines gives a mixed picture. For example, the various evaluation committees in Dutch science policy have taken a rather different attitude towards the inclusion of quantitative and bibliometric indicators (Van der Meulen 1992). The inclusion of citation data in the assessment of the performance of individual scientists is even more controversial. In science policy, it often amounts to a "not done", although citation data pop up in most evaluation exercises. The question of whether or not the emergence of the citation cycle has changed the notion of scientific quality in any fundamental way should be verified empirically[6].

Even if the concept of quality has not changed, however, the emergence of the citation cycle is a significant phenomenon in the scientific system. It constitutes an additional meta-information cycle. As a consequence, two analytically different science representations are produced. One is the domain of the expert in the

[5]Moed and Van Raan, Interview, Leiden 1995.

[6]Both the concept itself and the way it is used may vary and it may be difficult to distinguish between them.

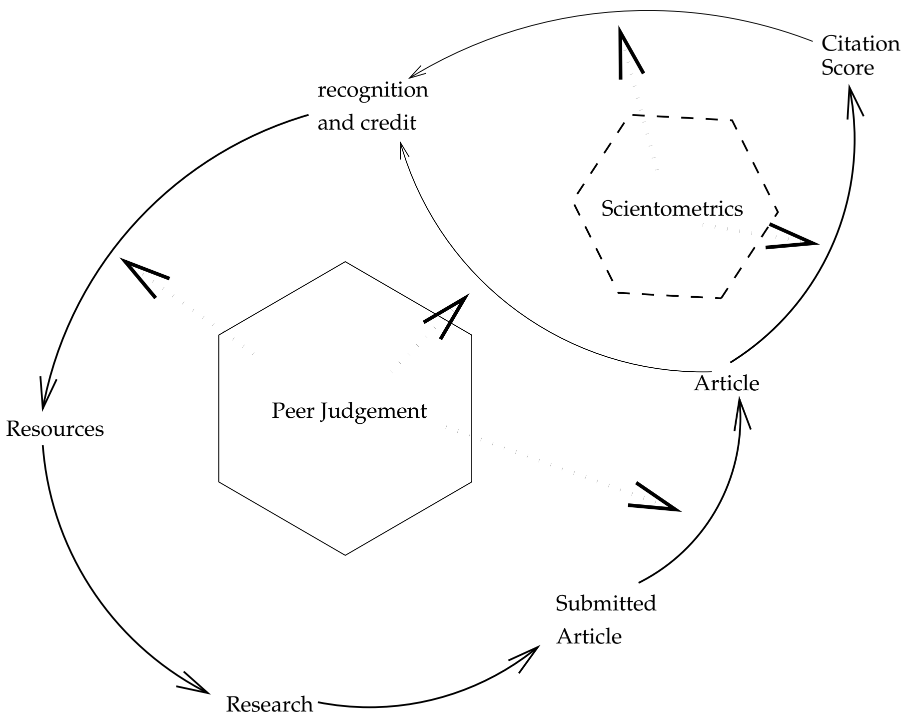

Figure 8.5: The adapted credibility cycle

field, the other the prerogative of the scientometrician. Since the regular evaluation of knowledge is a central axis around which the wheel of science spins, the citation cycle affects a vital part of scientific knowledge production. It would seem rather strange then, to dismiss science and technology indicators as being of only minor importance in the discussion on the research system's future. Admittedly, information cycles are rather elusive entities. Given their nature as hypothetical constructions, their existence cannot be proved or disproved directly. Nevertheless, the citation cycle has led to such material phenomena as controversies over the use of indicators, an institute like the Leiden Centre for Science and Technology Studies, the specialty of scientometrics, the journal *Scientometrics* and lately even an international society, the International Society for Scientometrics and Informetrics ISSI. In other words, the citation cycle has affected both the institutional structure of science and the discourse on science evaluation.

8.5 Paradigmatic versus formalized representations

8.5.1 Two representational domains

The citation culture has thus created two stable configurations: an intricate maze of science and technology indicators on the one hand, and the social science specialty of scientometrics on the other hand. Both have an important property in common: they represent science on the basis of formalized properties, be it in scientific literature, in its social or cognitive networks, or in its language. This is the main reason that the difference between the reference and the citation, made in chapter 1, is relevant to better understanding the dynamics of science representations in science as well as in science policy. In the sociology of science as well as in scientometrics, the resistance towards citation or co-word analysis has generally been interpreted in terms of vested interests. The same is true of the analysis of the resistance of scientometricians against critics of their practice. The question then is, however, where do these interests come from? Why do actors define them this way? And what cultural resources are mobilized by them? I would like to suggest that this is where the symbolic representations of science hinted at in chapter 1 come in.

Most of the time, science representations may not be very important. After all, researchers just do their job. A vague notion of the state of their specialty may be sufficient, even if it were inconsistent and half-baked. As soon as science itself is the topic of research, however, it does matter. Vague notions of technical experts then have to be translated into some explicit assessment of the state of the art in specialty X. All actors dealing with science and technology then have to more consciously manipulate representations of these phenomena. Practising scientists maintain a specific representation of their specialty, and of science in general, based on their reading of the literature, their expert discussions at conferences, and their appreciation of the skills of their peers. Engineers also base the representation of the technology they create on their specialized knowledge and know-how. In other words, their representations are based on the substance of the matter, on the substantive content of the scientific knowledge or on the specific skills embodied in the technology.

The scientometric representations of science and technology, whether embodied in indicators or in a scientometric article, are very different: they purposely abstract from the specific substance and are only based on the formalized relationships between the entities (scientists, publications, texts or artefacts). This difference, which all actors seem to notice in passing, is, I propose, the consequence of the fundamental difference in the concept of information that the various representations of science and technology build on. In other words, the two types of representations, produced in the peer review cycle and the citation cycle, not only differ with respect to their use but also with respect to their constitution.

8.5.2 Two concepts of information

Two different theoretical concepts of information have been used in the information sciences. The first was proposed by Shannon & Weaver (1949). These authors provided a theoretical underpinning of the encoding of information and thereby wished to solve the technical problem of noise on the communication channel. In their view, information is a countable entity. The amount of information enclosed within a certain message is equal to the average number of digits required to code it. This concept of information has nothing to do with meaning, it is an abstract, dimensionless entity, comparable with the thermodynamic entity entropy (because the form of the mathematical equations is the same). It is related to the paradigm of formal logic in mathematics:

> Logic, and by incorporation all of mathematics, was a game played with

> meaningless tokens according to certain purely syntactic rules. All meaning had been purged. One had a mechanical, though permissive (we would now say non-deterministic), system about which various things could be proved. Thus progress was first made by walking away from all that seemed relevant to meaning and human symbols. We could call this the stage of formalized symbol-manipulation. This general attitude is well reflected in the development of information theory. It was pointed out time and again that Shannon had defined a system that was useful only for communication and selection, and which had nothing to do with meaning. Regrets were expressed that such a general name as 'information theory' had been given to the field, and attempts were made to rechristen it as 'the theory of selective information' — to no avail, of course. (Newell & Simon 1990, 112)

A diametrically opposed concept of information was proposed by Bateson (1980). His concept puts central what Shannon & Weaver's (1949) filters away: meaning.

> In 1979 anthropologist-philosopher Gregory Bateson offered another definition of "information": *"Any difference which makes a difference."* He said, "The map is not the territory, we're told. Very well. What is it that gets from the territory to the map?" The cartographer draws in roads, rivers, elevations — things the map user is expected to care about. Data, signal ("news of a difference") isn't information until it means something or does something ("makes a difference"). The definition of information I kept hearing at the Media Lab was Bateson's highly subjective one. That's philosophically heartwarming, but it also turns out there's a powerful tool kit lurking in the redefinition. (Brand 1987, 78–79)

As has been discussed in chapter 6, for a few decades after World War II science was a self-governing domain. For this reason, the experts in specialty X still play a dominant role in judging the state of affairs in specialty X. They regularly produce an explicit representation of their specialty, an overall judgement which is based on the specialty-bound paradigm (Hoyningen-Huene 1993, 131–162). This paradigm is formed from an assessment of the meaning of the various bodies of knowledge, practices, institutions and so forth. Hence, it is based on the differences that make a difference within the specialty. This "paradigmatic science representation", to give it a name, is built of the stuff of Bateson's information concept. Evidently, this is true even if the specialty is Shannon's information science.

Since the seventies, science has increasingly been held accountable for its performance by science policy officials and the public at large. Sociologists of science, some historians of science and technology, as well as academically inclined science policy officials have been asked to study science in order to somehow steer science towards societal needs and desires. The study of specialty X by outsiders was thereby made possible and acceptable. Do these inquiries and representations from the outside occupy a different position? After all, historical and most sociological studies of science create different science representations from those that the specialists involved would produce. Yet, they also focus on meaning, on the difference that makes a difference. For example, the development of twentieth-century physics in Germany may be a story about the evolution of the Weimar republic. School formation in British astronomy may be explained in terms of, for example, differing styles of the leaders involved (Edge & Mulkay 1976). It is no longer the meaning of specialty X to itself, but meaning nevertheless reigns. It is no longer the paradigm of specialty X which dominates, but a different one, the historian's paradigm or the "disciplinary matrix" (Hoyningen-Huene 1993, 145) of the sociologist. Nevertheless, a paradigm does reign. These science representations from the outside are therefore also paradigmatic science representations based on meaning. They belong to the same class as the internal science representations discussed above[7].

This does not hold for the scientometric representation of science. Science and technology indicators create a "formalized representation" of science which initially neglects meaning. Of course, to interpret these representations one needs to attribute meaning again. The main point is, however, that this attribution of meaning can be postponed. This is crucial because it enables the manipulation of "meaningless" symbols, such as the citation. The sign citation is an entity like Shannon's information concept and like entropy. Dimensionless, meaningless, countable. The formalized science representation therefore does not use (or maybe better, is not informed by) Bateson's but by Shannon's information concept. If this line of reasoning is valid, the most important aspect of scientometrics may not be its numerical character. Porter (1995) has pointed to the objectifying role of numbers in bureaucratic and policy contexts (see also Van der Meulen 1992). This is also true of science and technology indicators. Nevertheless, it should be noted that not all scientometric representations are numerical in nature. Science maps in particular are geometric representations, although they are based on computations. Moreover, paradigmatic representations may also contain computations and numerical assessments, whether made by the field-specific expert or a historian or sociologist of science. Therefore, although scientometrics is a metrics and therefore numerical in character, this is not its most important feature. The distinction between the formalized and the paradigmatic representations is the more fundamental distinction.

This conclusion also reaches back to chapter 1: in itself the semiosis of the sign citation (chapter 1) is not important. After all, every citation can easily be converted back to its reference, and one can also create symmetrical tokens denoting interactive processes. This kind of re-translation is even an important part of the production process of the *SCI*[8]. The main reason that the semiosis of the citation and of the co-word are relevant, is the symbolic possibilities it creates to construct an encompassing formalized science representation that can compete at every level with paradigmatic science representations[9].

[7]Ashmore (1989) made the distinction between studying science from the inside or from the outside, an important one in his analysis. In the perspective of this study, this distinction seems, however, less important than the distinction based on the types of information used.

[8]Henry Small, discussion at the session on "The signs of science", 4S/EASST conference, New Orleans, 1995.

[9]This does not, in itself, explain why the formalized representations have become competitive. For this, one needs to look into the material interactions, i.e. the market for science indicators.

8.6 Indicator theories

The distinction between formalized and paradigmatic science representations can also be used to throw new light on the discussion on citation theories in the sociology of science and scientometrics. The search (in vain so far) for a definitive citation theory in scientometrics and the sociology of science has led to the regular production of reviews of theories of citation and citing (Narin 1976, Elkana, Lederberg, Merton, Thackray & Zuckerman 1979, Cozzens 1981, Cronin 1984, Edge 1977, Edge 1979, Gilbert & Woolgar 1974, Leydesdorff & Amsterdamska 1990, MacRoberts & MacRoberts 1989, Smith 1981). I will not repeat this here. The existing citation theories have tried to attribute meaning to the citation. Rephrased in more general terms, indicator theories try to attribute meaning to an individual indicator or set of indicators. In terms of the distinction between formalized and paradigmatic science representations, existing indicator theories in the sociology of science have grounded the meaning of an individual indicator (or set of indicators) in terms of a paradigmatic science representation. Using the criteria mentioned above, indicator theories can be grouped into four clusters, depending on the type of connection made between a set of indicators and a particular paradigmatic science representation: a "science of science" cluster, a "sociological" cluster, a "semiotic" cluster and an "information science" cluster of indicator theories. All theories put the relationship between a set of indicators and a specific paradigmatic representation in a central position.

- The science of science cluster

  The science of science cluster is represented first and foremost in Derek de Solla Price's work and in many publications in *Scientometrics*. The central notion is the reflexive use of science to measure itself (chapter 4):

  > I take the position that the workings of science in society show to a surprising degree the mechanistic and determinate qualities of science itself, and for this reason the quantitative social scientific investigation of science is rather more succesful and regular than other social scientific studies. It seems to me that one may have high hopes of an objective elucidation of the structure of the scientific research front, an automatic mapping of the fields in action, with their breakthroughs and their core researchers all evaluated and automatically signaled by citation analysis. (Price 1961, 194)

  This approach has not produced a very detailed, nor a consistent theory about the citation. It was assumed that the regularities of science as a social activity would more or less directly produce a well-structured set of data that could subsequently be used to access the hidden laws of the growth of science. Finding a simplest description of these laws, comparable with the way physicists go about their job, was the primary motive in this work. This also meant that playing with the data and using the citations rather loosely was perfectly permissible and even encouraged.

- The sociological cluster

  The science of science and the sociology of science both accepted citation frequency as a valid measure of scientific quality and as a sociometrically interesting link between authors or publications (chapter 4). The premise shared by everyone was that the reference and the citation were basically identical. This enabled an intuitive approach to the meaning of citation. Garfield, Price, Sher, and the sociologists of science at Columbia University all regularly referred to the act of citing to justify using citation data. For the latter group, this was especially appealing because the citation seemed to fit in very nicely with Merton's norms of science (Merton 1973). Norman Kaplan provided for the first explicit Mertonian explanation of the citation in an NSF-funded project (Kaplan 1965). In this perspective the citation is seen as the embodiment of the giving of recognition to which the scientist is obliged. Since this leads to a symmetrical positioning of the citation, it means that, provided the normal statistical precautions have been taken, the number of citations received is directly proportional to the recognition acquired.

  The Mertonian paradigm has produced a number of interpretations of the citation, each slightly different but all tied in with the central notion of science's specific norms and rules. In more general terms, Mertonian indicator theories try to explain the citation by relating it to the citing behaviour of the scientist or scholar. This has also been done by competing sociological paradigms. For example, citations have been interpreted as a form of persuasion (Gilbert 1977). This study has also used a sociological indicator theory, albeit implicitly. In chapter 7 the citation patterns have been used as sociometric indicators, providing information about the relationships in scientometrics at group level. The attempt to translate citation patterns into behavioural characteristics is the most common approach in citation theories. The sociological cluster is therefore the largest cluster of citation theories.

  The main reason that sociologists of science feel that this perspective has not produced the one encompassing citation theory, is the variety of behavioural characteristics underlying the citation patterns found in the literature. This is, however, the consequence of the semiotic inversion of the reference into the citation. This inversion is asymmetrical: whereas the references have very different characteristics (both textually and behaviourally), citations are all the same. The citation no longer betrays from what type of reference it was produced. This is why one should expect it to be difficult or even impossible to recreate the variety by citation analysis. Unless one re-translates the citation to the reference, that is, as is done in reference analysis. This is also why it is impossible to link the sign citation to some specific behavioural characteristic with respect to citing. Hence, although this type of research has delivered a reasonable amount of knowledge about citing cultures in science, in the quest for a citation theory it is a dead end.

- The semiotic cluster

In the semiotic cluster, created by the group of Michel Callon, not the citation but the co-word is the central "actant". It is, like the citation, made from a word (ususally either a keyword or a title word) by stripping every meaning from it. These co-words are subsequently used to construct maps that are claimed to be representative of fields of research and networks of relationships. This type of work was initially justified by criticizing the use of citation analysis, but in a thoroughly relativistic or a-sociological approach it has no need of this kind of justification. Different from the sociological cluster, these indicator theories do not try to relate to social behaviour but to semiotic networks. Intertextual relations between co-words are taken to reflect the development of networks of forces in science as well as the cognitive development of research. The signs live their own life and need no further justification.

- The information science cluster

  This cluster probably originated in the work of V. V. Nalimov in Russia (who saw science as a self-organizing process of information processing) in 1966, and ten years later in the work of Francis Narin (Narin 1976) in the United States, who took citation relations between journals as a sign of communication and the transmission of information. Parts of Garfield's writing can also be seen as the attempt from an information science point of view to the citation, as can some of Derek de Solla Price's articles. In general, the citation is related to processes of communication in science, or is seen as the upshot of an increasing complexity in these processes. Most of the time, the reference and the citation are taken to be identical. The distinctive feature of this group of theories is that it represents science as an information process, and abstracts itself from the substantive issues. In other words, it takes Shannon's information concept as point of departure for all of science.

All four types of approach ground the formalized framework of indicators in some paradigmatic science representation. Having a formalized paradigm, the third and fourth cluster of indicator theories transform all of science into a formal domain and obliterate the difference between the paradigmatic and the formalized representations. In contrast, what the first two clusters have in common is that they ignore the relationships between the indicators in the realm of the formalized science representation itself. In them, the paradigmatic domain is predominant.

8.7 The rise of the formalized

Chapter 1 argued that because of the difference between the reference and the citation, the legitimation of citation analysis should be analytically distinguished from the study of citing behaviour in science. The results of the subsequent chapters enable a more general formulation of this conclusion. Because of the emergence of the formalized representations, stimulated by the creation of the *SCI*, multiple relations have been created between the formalized and the paradigmatic representations of science (and technology). Every existing science or technology indicator theory is the embodiment of one possible type of relation within the domain of all possible relationships. Encompassing all this is not a sociological theory, but simply this proposal: to recognize the two different domains, to position each indicator theory accordingly, and to establish their interrelations.

In this sense, my proposal is also a theory, though a more abstract one: one could call this a proposal for a reflexive indicator theory. First, it is a theory about indicator theories because it explains how they can be related to one another and why the 30 year long quest for a citation theory has not been fruitful. Second, it is a theory about the indicators themselves, starting from the analytical distinction between the reference and the citation. Apparently, the two levels, usually kept strictly separate in science studies, go together seamlessly, an indication that the reflexivity issue can be fruitful indeed (Ashmore 1989).

As a theory about indicators, the reflexive citation theory borrows from all four clusters mentioned above. This proposal shares its reflexive character with the science of science cluster. The difference is in the appreciation of the nature of science. The sociological cluster has contributed to recognizing the variety in citing cultures upon which the reflexive citation theory builds further. The semiotic cluster and this proposal share the sensitivity to the way signs restructure and recreate reality. I have borrowed the use of Shannon's concept of information from the information science cluster. The difference with both the semiotic and the information science cluster is that these two clusters, each in their own way, attempt to translate the whole of science into their formalized domain because their paradigm is ultimately a formalized one. In the reflexive citation theory the playing field is more level: both concepts of information play an important role and are often combined in the creation of representations of science.

Until the 1960s, formalized representations of science were subsidiary to paradigmatic ones. The regular publication of the *SCI* since 1964 has enabled the citation representation of science to lead a more independent life. Combining citation data with other formalized data (e.g. coword data, econometric data, computerized full-text analysis, logfile analysis of documents in cyberspace) is presently enabling more complex formalized science representations[10]. It is now possible to combine various formalized and paradigmatic representations at the same time, for example to address policy relevant questions in more than one way. This is also the reason why different partial citation theories exist instead of one general one. It is both a matter of the analyst being able to adopt more than one perspective (Leydesdorff 1995), and the consequence of the relatively recent emergence of the domain of formalized representations in the scientific system.

Thus, the proposal for a reflexive indicator theory does not invalidate the use of specific indicator theories. It does, however, limit the claims to legitimation. The rift in science studies between scientometrics and more qualitative types of research is, I propose, not in the first place the reflection of the general cultural divide between number crunchers and innumerates. Rather it can be taken to represent, in the cognitive dimension, the difference between the formalized and

[10]Which, in itself, does not mean that they should also exist.

the paradigmatic science representations. The fact that each citation analysis has to justify itself afresh in each study is the consequence of the existence of multiple relationships between the two domains. If the formalized domain had been tightly connected to some paradigmatic domain, this might have been different because every citation analysis would have had at least initially a clearcut meaning[11].

The interaction between formalized and paradigmatic science representations also limit the extent to which the formalized domain might "take over". First of all, people tend to attribute meaning to quantitative indicators or analyses. If these must circulate in society, they will therefore have to be translated into some paradigm. Second, the extent to which the formalized representations can alter the production of scientific knowledge or technological know-how is restricted by the embodiment of these practices. For example, it is not likely that publishing in science could completely be corrupted into a practice of collecting citations as a means in itself. Neither is it likely that the scientometric maps of science could replace the writing of reviews. Thus, if the formalized representations play any role, it will be more in the form of hybrid representations of science and technology than as "pure" formalized ones. Scientists, engineers and science policy officials will therefore have to handle both paradigmatic and formalized representations. Citing and publication patterns may be developed with an eye to both domains. The importance attributed to journals that are covered in the production of the *SCI* ("ISI journals"), is an indication of this trend.

The emerging new regime of electronic scientific publication may moreover impose a new relationship upon the paradigmatic and formalized science representations. "Citation" as well as "co-word" may acquire a novel function, at all levels. Since electronic publication seems able to produce and process a larger number of formal records, formalized representations can be expected to play an increasing role, both as an information retrieval tool and as forms of on-line quality control (crucial in an era of "accountable science"). Formalized representations may also increasingly appear in distributed form. For example, whereas the *SCI* had to be published separately from its sources and references, electronic publishing enables "live citation indexes" as part of the cited publications. Formalized representations may also be embodied in simulations of science. Being essentially algorithmic, the formalized domain may also become a point of departure for science (and scientometrics) by robots. It may for example become feasible to have computers construct new indicators in an evolutionary simulation of the scientific and technological system.

In short, the rise of a variety of formalized science representations, as well as an increasingly intimate interaction between formalized and paradigmatic science representations, may be a lasting result of the emergence of the citation culture in science.

Back to Contents Page

[11] Although it would not have prevented the possibility of several interpretations.

Back to Contents Page

Samenvatting

Wetenschappers moeten in toenemende mate verantwoording afleggen van hun handelen. Deze ontwikkeling heeft een aantal nieuwe beroepen in het leven geroepen, waaronder dat van de sciëntometrist. Deze expert legt de wetenschap wetenschappelijk de maat, doorgaans ten behoeve van een of andere vorm van wetenschapsbeleid. Sciëntometristen zijn gespecialiseerd in het beoordelen en in kaart brengen van wetenschap met behulp van uitgebreide gegevensbestanden over de wetenschappelijke literatuur. In deze databanken, de *Science Citation Index*, de *Social Science Citation Index* en de *Arts & Humanities Citation Index* wordt niet zozeer de inhoud van publicaties weergegeven, alswel hun formele, bibliografische kenmerken zoals titel, naam van de auteurs, het aantal referenties en niet te vergeten de citaties die aan de publicaties zijn toegekend.

De sciëntometrie is niet alleen een beleidsinstrument maar ook een sociale wetenschap. Het vakgebied kent een centraal tijdschrift, *Scientometrics* en er worden regelmatig internationale conferenties gehouden georganiseerd door de International Society for Scientometrics and Infometrics. Er bestaan op dit moment een paar honderd sciëntometristen in de wereld. Deze studie betoogt dat de ontwikkeling van dit vakgebied vruchtbaar kan worden geanalyseerd als neerslag en indicator van een nieuwe subcultuur in de wetenschap: *de citatiecultuur*. Deze subcultuur heeft vrijwel ongemerkt de essentie van wetenschapsbeleid getransformeerd. In dit proefschrift probeer ik deze verandering te onderkennen en haar betekenis voor de kennisproductie in kaart te brengen.

Wetenschappelijke artikelen zijn onder andere herkenbaar aan hun referenties naar ander wetenschappelijk werk. Deze literaire gewoonte is aan het eind van de vorige eeuw ontstaan en is heden ten dage een herkenbare karakteristiek van wetenschappelijk werk geworden. In het eerste hoofdstuk wordt aangegeven dat de precieze stijl van refereren gebiedsafhankelijk is, waardoor we in feite kunnen spreken van verschillende citeerculturen in de wetenschap. Het tot bloei komen van deze citeerculturen schiep een onvoorziene bron van gegevens voor onderzoek en beleid: citatiegegevens. Het lijkt achteraf gezien vrijwel onvermijdelijk. Onderzoekers citeren immers het werk van collega's als zij dat nuttig hebben bevonden. Die publicatie heeft blijkbaar meer nut dan een niet-geciteerde. Het aantal malen dat een artikel wordt geciteerd lijkt dan ook niet alleen een voor de hand liggende, maar ook een accurate maatstaf van wetenschappelijke invloed of kwaliteit. Deze redenering gaat er van uit dat de citatie index iets meet dat voordien al bestond. Is dat echter wel zo voor de hand liggend?

Om hierover meer te weten te komen analyseert deze studie de citatiecultuur op basis van het analytische onderscheid tussen de citatie en de referentie. Hoofdstuk 1 beargumenteert, geïnspireerd door de semiotiek (tekenleer), waarom het zinnig is de referentie en de citatie in hun kwaliteit van teken te behandelen. De zaken komen dan in een ander licht te staan. Aangezien beide tekens elk een verschillende referent hebben zijn ze niet identiek maar elkaars spiegelbeeld. Toch worden ze in de wetenschapssociologie in het algemeen als identiek behandeld. Als we deze, vaak impliciete, vooronderstelling laten varen blijkt de citatie niet zozeer het product van de wetenschappelijk onderzoeker te zijn maar veeleer dat van de producent van de citatie-index. Vandaar dat dit onderzoek begint bij de schepping van de *SCI* door Eugene Garfield in de jaren zestig in de Amerikaanse stad Philadelphia.

De twee volgende hoofdstukken behandelen het ontstaan van de *SCI* op basis van archiefonderzoek van de originele documenten en briefwisselingen tussen de belang rijkste actoren. De oorsprong van de citatie-index ligt niet in de wetenschappelijke wereld maar in die van de Amerikaanse rechtspraak. De scheppers van de *SCI* blijken niet alleen een bibliografisch instrument te hebben gebouwd, maar tegelijkertijd een symbolische vertaling van een juridisch element in een nieuw wetenschapsteken te hebben bewerkstelligd. Hun onderneming blijkt bovendien onverwacht verweven met cruciale debatten over wetenschapsbeleid en politiek in de VS.

De volgende hoofdstukken pogen de consequenties van het ontstaan van de *SCI* in 1964 in kaart te brengen. Het vierde hoofdstuk verhaalt over de receptie van de *SCI* door wetenschapssociologen enerzijds en de deelnemers aan de "wetenschap van de wetenschap" anderzijds. Beide groepen reageerden verschillend. Het hoofdstuk suggereert dat de huidige sciëntometrie kan worden gezien als de samensmelting van de opkomende citatiecultuur en de oudere traditie van de "wetenschap van de wetenschap".

Wellicht nog belang rijker is het geheel aan wetenschaps- en technologie-indicatoren dat is ontstaan op basis van de citatie-indexen. Dit is het onderwerp van het vijfde hoofdstuk dat geen historische, maar een conceptuele analyse is van het tekensysteem dat op grondslag van de *SCI* is opgetrokken. De belangrijkste nu in gebruik zijnde indicatoren worden onder de loep genomen. Het hoofdstuk laat zien hoe het onderscheid uit het eerste hoofdstuk tussen referentie en citatie vruchtbaar gemaakt kan worden voor de deconstructie van deze indicatoren. Waar indicatoren in het algemeen beoordeeld worden op hun consistentie blijkt de creatie van sciëntometrische indicatoren een verre van eenduidig proces. Hoewel dit hun gebruik in beleid zeker niet ontkracht, werpt het wel interessant licht op hun karakter als semiotische en socio-cognitieve producten van creatieve sciëntometristen.

Het creatief omgaan met indicatoren is ook een kenmerk van wetenschapsbeleid. In het ontstaan van de sciëntometrie en zijn indicatoren heeft de beleidsbehoefte aan meetinstrumenten een belang rijke rol gespeeld. Zonder deze "trekkracht" (in onderscheid met de "duwkracht" die het nieuwe tekensysteem zelf teweegbrengt) zou de sciëntometrie niet zijn huidige vorm hebben gekregen. Hoofdstuk 6 behandelt op basis van historisch archiefonderzoek en interviews het ontstaan van de markt voor wetenschaps- en technologie-indicatoren in Nederland, een proces waarin de Raad van Advies voor het Wetenschapsbeleid (RAWB) een katalyserende rol heeft gespeeld.

Nadat het zevende hoofdstuk een reflexief kwantitatief sciëntometrisch portret heeft pogen te schetsen, gaat het slothoofdstuk in op de mogelijke implicaties van de bevindingen van dit proefschrift. Het laat zien dat het geheel van wetenschaps- en technologie-indicatoren een nieuwe representatie van wetenschap tot stand heeft gebracht. Als we kennisproductie analyseren als een cyclisch proces waarin informatie wordt geschapen en uitgewisseld, verschijnen de nieuwe indicatoren als een tweede cyclus, gekoppeld aan de primaire. Waar in de primaire cyclus de inhoud en betekenis van de informatie centraal staat, is de secundaire cyclus het domain van de formalisering. Dit valt samen met het klassieke onderscheid in definities van het begrip informatie. De discussie over het gebruik van de citatie-frequentie en afgeleide indicatoren stoelt zo bezien op de vertaling van patronen in het ene domein naar het andere. Het nieuwe formele type representatie van wetenschap is ook de oorzaak dat de stand van zaken in de wetenschap door twee verschillende, en elkaar deels beconcurrerende typen expertise kan worden beoordeeld, die van de vakgenoten en die van de sciëntometristen.

Dit hoofdstuk laat bovendien zien dat de in deze studie gehanteerde benadering ook een nieuwe formulering van een oud probleem in de wetenschapssociologie met zich meebrengt. Sinds ruim dertig jaar geleden de strijd rondom de citatie-analyse ontbrandde, hebben sociologen en sciëntometristen gepoogd de grondslag van citatie-analyse te vinden in het citeergedrag van wetenschappers. Tot nog toe tevergeefs. Deze studie draagt een verklaring aan voor dit falen en stelt bovendien een nieuw, reflexief type citatietheorie voor.

Back to Contents Page

Back to Contents Page

ISI Press Release

1961

NEWS RELEASE

Contact: Mrs. Joan E. Shook

INSTITUTE FOR SCIENTIFIC INFORMATION 33 SOUTH SEVENTEEN STREET PHILADELPHIA 3, PA.
phone/locust 4-4400 cable/currcon twx/ph 803

For Immediate Release

$300,000 GRANT TO PROBE INFORMATION RETRIEVAL AWARDED TO INSTITUTE FOR SCIENTIFIC INFORMATION BY NATIONAL INSTITUTES OF HEALTH AND NATIONAL SCIENCE FOUNDATION... THREE YEAR PROJECT TACKLES CITATION INDEX TECHNIQUES FOR SCIENCE

Research scientists will soon be consulting a more precise and specific literature index that links together subject material that would never be collated by usual indexing systems. Concerned with new starting points for scientific literature searches, the unique concept uncovers sometime-buried associations, relating important works and authors, yet keeps the researcher abreast of the masses of current published scientific information. This new approach to information retrieval is called the Citation Index.

A $300,000 grant extending over a three-year period has been awarded to the Institute for Scientific Information, Philadelphia, Pennsylvania, to study the practicability of citation indexes and to test their techniques of preparation. The project, under joint sponsorship of the National Institutes of Health and the National Science Foundation, is aimed at producing a unified citation index for science including the publication of a genetics index.

Dr. Eugene Garfield, director of ISI, explains the simplified citation index this way. If this article you are now reading were processed for citation indexing, it would be called the ‘‘referant’’. All the items in the bibliography would be called ‘‘references’’. A list of these references, each of which is followed by its associated list of referants, becomes the citation index. By using a citation index, the researcher finds out what works have cited a particular reference following its publication.

By focussing on the individual citation rather than specific subjects, the citation index signifies a more sophisticated method of scientific documentation, as well as a growing bibliographical aid. Because the scientific researcher is generally aware of one or more particular papers already published in the area of his specialized interests, he will use the citation index as a starting point, rather than the specific topics found in conventional indexing.

Better scientist-to-scientist communication is expected, for authors can see at a glance what literature has been published since their works, in their own and related fields, that refer back to their own works.

By virtue of its different construction and purpose, the citation index complements indexes like Beilstein, Chemical Abstracts, Biological Abstracts and is not intended as a substitute.

''Experimental studies on systems for extracting and processing citations will soon be completed'', Dr. Garfield reports. It is estimated that approximately one million references will be processed during the next six months on a high speed magnetic tape computer. According to Dr. Garfield, ''approximately three million citations, from scientific literature published in the five-year period 1959 to 1963, will be processed during the life of the project.''

As a guide to the project and aid towards refining concepts of the methodology, an advisory committee of scientists has been established. Members of this board are Dr. Gordon Allen, National Institute of Mental Health; Dr. Joshua Lederberg, Stanford University; Dr. George LeFevre, Harvard University; Dr. Joseph Melnick, Baylor University; Dr. Sol Spiegelman, University of Illinois, Dr Luca Cavalli-Sforza, Institute di Genetica, Pavia.[12]

Back to Contents Page

[12]As published in ISI (1961). An undated version of this press release without the name of Dr. Cavalli-Sforza is present in Garfield's Personal Archive in Philadelphia. The exact date the press release was sent out is unknown to me. It must have been on or shortly after 17 May 1961, because on that day Garfield received the signed contract from NSF (Garfield to Lederberg May 17, 1961). On 26 July 1961 Lederberg suggested Cavalli-Sforza as foreign geneticist on the Board (Lederberg to Garfield, July 26, 1961). Later Victor A. McKusick, M.D., Johns Hopkins University joined the advisory board (Garfield & Sher 1963, vi).

Back to Contents Page

The Weinberg report on citation indexing

M. ***Citation Indexing Should Be Useful***

Along with development of hardware, much ingenious thought must obviously go into software; i.e., indexing and other preparation of the documents for subsequent retrieval. Of the new approaches to software, the Panel is particularly impressed with the citation index; we wish to call the technical community's attention to this apparently powerful, though relatively little used, new searching tool.

All of us are familiar with lists of references at the end of an article. Such lists enable the reader to trace backward in time the antecedents of the article being perused. Every scientist has used such lists to delve more deeply into the subject he is studying. But reference lists only go backward in time; they give no hints as to the influence a given article has had on the development of the subject after the article appeared in print. The citation index is a list of the articles that, subsequent to the appearance of an original article, refer to or cite that article. It enables one to trace forward in time the same sort of interconnections with the literature that, by means of lists of references, one now traces backward in time. Because the indexing is based on the author's, rather than on the indexer's, estimate of what articles are related to what other articles, citation indexes are particularly responsive to the user's, rather than to the indexer's, viewpoint.

Lawyers have used a citation index, Shepard's Citations, for more than 100 years. Each year Shepard's lists all appellate decisions that have cited any previous cases. Since the law is unified in somwhat the same way as is science in that the rule of precedent connects what happens later with what happened earlier, it is not surprising that a bibliographic tool so useful to the lawyer could also be useful to the scientist.

The National Science Foundation is sponsoring trials of citation indexing in genetics and in statistics and probability. The genetics index, for example, will cover all the genetics literature from 1959 through 1963 and will be published in a singel volume; it will be kept up to date by yearly supplements. The Panel believes that citation indexing, particularly in combination with permuted title indexing, will come to be used widely, and that its use will further alter both the way in which we think of the technical literature and the way we manage it.

Back to Contents Page

Back to Contents Page

Note on archives and interviews

.1 Archives

The following archives were studied:

- Eugene Garfield's personal archive, ISI, Philadelphia
- The RAWB archive, The Hague
- NSF Historian's archive, Washington DC
- The Derek Price personal archive, La Villette, Paris
- The FOM archive, Utrecht
- The Ministry of Education HOW archive, The Hague
- Ben Martin's personal archive, SPRU, Brighton
- Selected items from Francis Narin's archive, CHI, New Jersey
- A few selected items from Tibor Braun's archive, Budapest
- Selected items from the CWTS archive, Leiden
- Library of Congress, Washington DC
- Hildrun Kretschmer's personal archive, Berlin

.2 Interviews

The following persons have been interviewed in the course of the research reported on in this thesis:

Beverly Bartolomeo, Donald D. de Beaver, Manfred Bonitz, Tibor Braun, Emiel Broesterhuizen, Michel Callon, Stephen Cole, Jean-Pierre Courtial, Bob Coward, Suzan Cozzens, Leo Egghe, Helen Gee, Eugene Garfield, Michael Gibbons, Wolfgang Glänzel, Isabelle Gomez and her colleagues, Arie van Heeringen, Diana Hicks, Wim Hutter, Phoebe Isard, Sheila Jasanoff, Sylvan Katz, Mike Koenig, Hildrun Kretschmer, Bruno Latour, Joshua Lederberg, Cees le Pair, Terttu Luukkonen, Morton Malin, Ben Martin, Robert King Merton, Henk Moed, Francis Narin and his staff, Ton Nederhof, Ton van Raan, Henk Rigter, Arie Rip, Jo Ritzen, Ronald Rousseau, András Schubert, Irving Sher, Len Simon, Henry Small, Jan van Steen, Peter Tindemans, William Turner, John Ziman, Harriet Zuckerman

Together they have provided much more material than I could possibly include in this thesis.

Back to Contents Page

Listings of PERL software used in this thesis

.1

```
#!/usr/bin/perl

# COUNT CITATION FREQUENCY AUTHORS IN SCIENTOMETRICS

# OPEN AUTHOR NAME DATABASES

open(IDINPUT, "<Authors.ID.Input.txt");
@idinput = <IDINPUT>;
chop;

for (@idinput) {
($authorin,$idin) = split(/: /);
$id =˜ s/\s$//; # remove trailing spaces
$IDIN{$authorin} .= $idin;
}

open(IDOUTPUT, "<Authors.ID.Output.txt");
@idoutput = <IDOUTPUT>;
chop;

for (@idoutput) {
($idout,$authorout) = split(/: /);
$authorout =˜ s/\s$//; # remove trailing spaces
$IDOUT{$idout} .= $authorout;
}

# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;
```

```
# COUNT CITATION FREQUENCY

# PROCESS EACH RECORD

for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);
    @cr = grep(s/^CR-\s//,@fields);

# TRANSLATE ARRAY VALUES IN STRING VALUES

    $cr = shift @cr;
    $cr =~ s/^\n//;

# SPLIT FIELDS AND COUNT

    @reference = split(/^/,$cr);

REF: { for (@reference) {

($citedauthor, $rest) = /^\s*(\b\D*)(,\s.*)/;
&IDINPUT;
chop($citedauthor); # remove added newline in subroutine
&IDOUTPUT;
$CitatFreq{$citedauthor}++;

}
}
}

#PRINT RESULTS

open(CITFREQ, ">Citation_Frequency");
print CITFREQ "Citation Frequency Cited Authors";

open(CITFREQ, ">>Citation_Frequency");

foreach (sort keys(%CitatFreq)) {

    print CITFREQ ($_, " ", $CitatFreq{$_});

# Prepare ranking

$CitatRank{$CitatFreq{$_}} .= $_ . " ";

# Prepare computation distribution

$DistCit{$CitatFreq{$_}}++;

}

open(CITRANK, ">Citation_Frequency_Ranked");
```

```
print CITRANK "Ranked Citation Frequency Cited Authors";

open(CITRANK, ">>Citation_Frequency_Ranked");

foreach (reverse sort bynumber keys(%CitatRank)) {

print CITRANK ($_, " ", $CitatRank{$_});

}

open(DISTCIT, ">Citation_Frequency_Distribution");

print DISTCIT "Distribution of citation frequency: citations, number
of cited authors";
open(DISTCIT, ">>Citation_Frequency_Distribution");

foreach (sort bynumber keys(%DistCit)) {

print DISTCIT ($_, " ", $DistCit{$_});

}

#SUBROUTINES

sub bynumber { $a <=> $b;}

sub IDINPUT {
if ($IDIN{$citedauthor}==0) {
  next REF;
}
else {
$citedauthor = $IDIN{$citedauthor};
}
}

sub IDOUTPUT {
$idout = $citedauthor;
$citedauthor = $IDOUT{$idout};
}
```

.2

```
#!/usr/bin/perl

# UNIFY CITING AND CITED AUTHOR NAMES IN SCIENTOMETRICS


# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
```

```
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "<../data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;

# PROCESS EACH RECORD

for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);

# AUTHOR FIELD

    @au = grep(s/^\n*AU-\s//,@fields);

    unless (@au == ()) { # als geen au dan verder met references

    $au = shift(@au);

@author = split(/;/,$au); # extract citing author names

grep (s/^\s*//,@author); # remove leading spaces
grep (s/\s*$//,@author); # remove trailing spaces
grep (s/^\n*//,@author); # remove leading newlines
grep (s/\n*$//,@author); # remove trailing newlines



# BUILD ARRAYS

    foreach $i (@author) {

if ($AUTHORID{$i} == "") {

    $counter++;
    $AUTHORID{$i} = $counter;


}
    }
}

# REFERENCE FIELDS

   @cr = grep(s/^\n*CR-\s//,@fields);

    if (@cr == ()) {next;}

    $cr = shift(@cr);
```

```
# SPLIT CR-FIELD INTO INDIVIDUAL REFERENCES

    @reference = split(/^/,$cr);
grep (s/^\s*//,@reference); # remove leading spaces
grep (s/\s*$//,@reference); # remove trailing spaces
grep (s/^\n*//,@reference); # remove leading newlines
grep (s/\n*$//,@reference); # remove trailing newlines


grep ($_ =~ s/^(\w+\s*\w*)\W(.*)/\1/,@reference); # extract cited authors



    foreach $i (@reference) {

if ($AUTHORID{$i} == "") {

    $counter++;
    $AUTHORID{$i} = $counter;


}
    }
}



#PRINT THE RESULTS

open(AUTHORID, ">../Author.ID");
print AUTHORID "Author ID's Input File:";
open(AUTHORID, ">>../Author.ID");


foreach (sort keys(%AUTHORID)) {

    print AUTHORID ($_, ": ", $AUTHORID{$_});
}


sub bynumber { $a <=> $b; }
```

.3

```
#!/usr/bin/perl

$\ = "\n";

# Use cleaned up author names to recompute productivity

open(AUTHORS, "<Publications_author.txt") || die "Can't open author
```

```
file: $!\n";

@authors = <AUTHORS>;

for (@authors) {

    s/(\s\d+)\s*\n$/:$1/;
    ($name, $pubnumber) = split(/:/);

    $productivity{$pubnumber}++;

}

# Print results

open(PRODUCT, ">Productivity_authors.txt");
print PRODUCT "";
open(PRODUCT, ">>Productivity_authors.txt");

foreach $pubnumber (sort bynumber keys(%productivity)) {

    print PRODUCT ($pubnumber, " ", $productivity{$pubnumber});
}

sub bynumber { $a <=> $b}
```

.4

```
#!/usr/bin/perl

# Print output default newline

$\ = "\n";

# CREATE PURIFIED DATABASE OF UNIQUE AUTHOR NAMES FOR PRINT OUTPUT

# FIRST OPEN DATABASE OF UNIQUE AUTHOR NAMES

open(ID, "<Authors.ID.Input.txt") || die "Can’t open: $!\n";

@ID = <ID>;
shift; # remove first line

for (@ID) {

    chop; # remove newline
($author,$id) = split(/: /);
$id =˜ s/\s$//; # remove trailing spaces
if ($id == 0) {next;}
$AUTHORID{$id} .= $author . ",";
```

```
}

# CHECK TO SEE WHETHER NAME IS UNIQUE (CHECK ON COMMA)

foreach (sort keys(%AUTHORID)) {

    chop $AUTHORID{$_}; # remove trailing comma


if (($n = index($AUTHORID{$_},",")) >= 0) {


    $AUTHORID{$_} = substr($AUTHORID{$_},0,$n);
}


# CREATE OUTPUT DATABASE

$AUTHORPURE{$_} .= $AUTHORID{$_};

}

# PRINT RESULTS

open(PURIFY, ">Authors.ID.Output.txt");
print PURIFY "Author ID’s purified:";

open(PURIFY, ">>Authors.ID.Output.txt");

foreach (sort bynumber keys(%AUTHORPURE)) {

    print PURIFY ($_, ": ", $AUTHORPURE{$_});
}

sub bynumber {$a <=> $b;}
```

.5

```
#!/usr/bin/perl

# OPEN AUTHOR NAME DATABASES

open(IDINPUT, "<../Authors.ID.Input.txt") || die "Can’t open
Authors.ID.Input.txt: $!\n";

@idinput = <IDINPUT>;
chop;

for (@idinput) {
($authorin,$idin) = split(/: /);
$id =˜ s/\s$//; # remove trailing spaces
$IDIN{$authorin} .= $idin;
```

```
}

open(IDOUTPUT, "<../Authors.ID.Output.txt") || die "Can’t open
Authors.ID.Output.txt: $!\n";

@idoutput = <IDOUTPUT>;
chop;

for (@idoutput) {
($idout,$authorout) = split(/: /);
$authorout =~ s/\s$//; # remove trailing spaces
$IDOUT{$idout} .= $authorout;
}


# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "<../data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;

# PROCESS EACH RECORD

for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);
@au = grep(s/^AU-\s//,@fields);
$au = shift @au;
$au =~ s/^\n//; # remove leading newline

@authors = split(/;/,$au);
grep(s/^\s*//,@authors); # remove leading spaces

AUT: {foreach $author (@authors) {

    if (!defined($author)) {next;}
    &IDINPUT;
    if (($author == 0) || (!($author))) {next;}
    &IDOUTPUT;
    $productivity{$author}++;
}
  }
}

# COMPUTE DISTRIBUTION PRODUCTIVITY OVER AUTHORS

foreach $author (sort keys(%productivity)) {
```

```
    $proddist{$productivity{$author}}++;
}


# RANK AUTHORS ACCORDING TO PRODUCTIVITY

foreach $author (sort keys(%productivity)) {
    $prodrank{$productivity{$author}} .= $author . ",";
}

# Print results

open(PRODUCT, ">Productivity_authors.txt");
print PRODUCT "Number of publications per author";
open(PRODUCT, ">>Productivity_authors.txt");

foreach (sort bynumber keys(%productivity)) {

    print PRODUCT ($_, " ", $productivity{$_});
}

open(PRODDIST,">Productivity_authors.Distribution.txt");
print PRODDIST "Distribution of productivity over authors";
open(PRODDIST,">>Productivity_authors.Distribution.txt");

foreach (sort bynumber keys(%proddist)) {

    print PRODDIST ($_, " ", $proddist{$_});
}

open(PRODRANK,">Productivity_authors.Ranked.txt");
print PRODRANK "Number of publications per author ranked";
open(PRODRANK,">>Productivity_authors.Ranked.txt");

foreach (reverse sort bynumber keys(%prodrank)) {

    chop($prodrank{$_});
 if ($_ <= 1) {next;}
    print PRODRANK ($_, " ", $prodrank{$_});
}



# SUBROUTINES

sub bynumber { $a <=> $b}

sub IDINPUT {

$author = $IDIN{$author};
chop($author);

}

sub IDOUTPUT {
```

```
$author = $IDOUT{$author};
}
```

.6

```
#!/usr/bin/perl

# Program to compute overlap between two data arrays,
# Scientometrics


# Ensure paragraph mode again

$/ = "";
$* = 1;

# Open arrays

open(SCIENTOMETRICS, "<data.unix.txt");
open(SCIENTO2, "<data2.txt");

# Localize temp array

local(%mark);

@records = <SCIENTOMETRICS>;
@records2 = <SCIENTO2>;

# Compute overlap

grep($mark{$_}++,@records);

@result = grep($mark{$_},@records2);

print @result;
```

.7

```
#!/usr/bin/perl
# Program to clean Dialog DATA on Scientometrics of funny characters

# Enable paragraph mode

$/ = "";
$* = 1;

# Open the database, complain if impossible
```

```
open(SCIENTOMETRICS, "<data") || die "Can't open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;

# Strip all elements of funny characters

for (@scientometrics) {

s/\cC|\cZ|\cD|\cW|\cA|\cB|\cE//g;

}

open(CLEANED, ">data.clean") || die "Can't open data.clean: $!\n";

$\="\n";
print CLEANED @scientometrics;
```

.8

```
#!/usr/bin/perl

# MEASURE CITATION-RELATIONSHIPS IN SCIENTOMETRICS
# FOR FURTHER STRUCTURAL ANALYSIS (STRUCTURE)

# OPEN AUTHOR NAME DATABASES

open(IDINPUT, "<../Authors.ID.Input.txt");
@idinput = <IDINPUT>;
pop;

for (@idinput) {
($authorin,$idin) = split(/: /);
$idin =˜ s/\s*$//; # remove trailing spaces
$idin =˜ s/\n*$//; # and newlines
$authorin =˜ s/\s$//; # remove leading spaces
$authorin =˜ s/\s$//; # and newlines


$IDIN{$authorin} = $idin;
}

# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "../data") || die "Can't open data: $!\n";
```

```
@scientometrics = <SCIENTOMETRICS>;

# CREATE ASSOCIATIVE ARRAY OF PRODUCTIVITY BY ID’S

for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);

#AUTHOR FIELDS

    @au = grep($_ =~ s/^\n*AU-\s//,@fields);
    if (@au == ()) {next;}
    $au = shift(@au);

@author = split(/;/,$au); # extract citing author names

grep (s/^\s*//,@author); # remove leading spaces
grep (s/\s*$//,@author); # remove trailing spaces
grep (s/^\n*//,@author); # remove leading newlines
grep (s/\n*$//,@author); # remove trailing newlines

grep ($_=$IDIN{$_},@author); # replace authornames with ids


foreach $i (@author) {

    unless ($i == -1 || $i == 0) { # do not include non-authors
$product_byid{$i}++;
}
}

}


# BUILD ASSOCIATIVE ARRAYS CITING CITED RELATIONSHIPS

# OPEN THE DATABASE AGAIN

open(SCIENTOMETRICS, "../data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;


for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);

#AUTHOR FIELDS

    @au = grep($_ =~ s/^\n*AU-\s//,@fields);
    if (@au == ()) {next;}
```

```
    $au = shift(@au);

@author = split(/;/,$au); # extract citing author names

grep (s/^\s*//,@author); # remove leading spaces
grep (s/\s*$//,@author); # remove trailing spaces
grep (s/^\n*//,@author); # remove leading newlines
grep (s/\n*$//,@author); # remove trailing newlines

grep ($_=$IDIN{$_},@author); # replace authornames with ids



# REFERENCE FIELDS

    @cr = grep(s/\n*^CR-\s//,@fields);
    if (@cr == ()) {next;}

    $cr = shift(@cr);

# SPLIT CR-FIELD INTO INDIVIDUAL REFERENCES

    @reference = split(/^/,$cr);
grep (s/^\s*//,@reference); # remove leading spaces
grep (s/\s*$//,@reference); # remove trailing spaces
grep (s/^\n*//,@reference); # remove leading newlines
grep (s/\n*$//,@reference); # remove trailing newlines


grep ($_ =~ s/^(\w+\s*\w*)\W(.*)/\1/,@reference); # extract cited authors

grep ($_=$IDIN{$_},@reference); # replace cited authornames with ids

# CREATE ASSOCIATIVE CITATION ARRAY FOR PROD > 1

foreach $i (@author) {

    if ($product_byid{$i} <= 1) { next;}

    foreach $j (@reference) {

if ($product_byid{$j} <= 1) { next;}

    $citationid = $i . "-" . $j;
    $citation{$citationid}++;
    }
}
}


# CREATE NORMAL ARRAY AND TURN THIS INTO STRING

foreach $n (sort bynumber keys %product_byid) {

    @finalstring = ();
```

```
    if ($product_byid{$n} <= 1) {next;}

foreach $m (sort bynumber keys %product_byid) {

    if ($product_byid{$m} <= 1) {next;}

    $citationid = $n . "-" . $m;

    if ($citation{$citationid} == "") {

$citation{$citationid}= "0"; # purely for presentation
    }

    push(@finalstring,$citation{$citationid});
}

$aantal = @finalstring;
$citedauthor_structure_matrix{$n} = pack("A3" x $aantal,@finalstring);

}

# PRINT FINAL CITED AUTHOR MATRIX FOR STRUCTURE

    open (STRUCTUREMATRIX,">../Cited_authors_Structure_Matrix");
print STRUCTUREMATRIX ("Cited authorship matrix for Structure");
    open (STRUCTUREMATRIX,">>../Cited_authors_Structure_Matrix");

foreach (sort bynumber keys %citedauthor_structure_matrix) {

    print STRUCTUREMATRIX $citedauthor_structure_matrix{$_};
}


$n = keys(%citedauthor_structure_matrix);
print $n;



sub bynumber { $a <=> $b; }
```

.9

```
#!/usr/bin/perl

# MEASURE CO-AUTHORSHIPS IN SCIENTOMETRICS

# OPEN AUTHOR NAME DATABASES


open(IDINPUT, "<../Authors.ID.Input.txt");
@idinput = <IDINPUT>;
```

```
pop;

for (@idinput) {
($authorin,$idin) = split(/: /);
$idin =˜ s/\s*$//; # remove trailing spaces
$idin =˜ s/\n*$//; # and newlines
$authorin =˜ s/\s$//; # remove leading spaces
$authorin =˜ s/\s$//; # and newlines


$IDIN{$authorin} = $idin;
}

# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "../data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;

# CREATE ASSOCIATIVE ARRAY OF PRODUCTIVITY BY ID’S

for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);

#AUTHOR FIELDS

    @au = grep($_ =˜ s/ˆ\n*AU-\s//,@fields);
    if (@au == ()) {next;}
    $au = shift(@au);

@author = split(/;/,$au); # extract citing author names

grep (s/ˆ\s*//,@author); # remove leading spaces
grep (s/\s*$//,@author); # remove trailing spaces
grep (s/ˆ\n*//,@author); # remove leading newlines
grep (s/\n*$//,@author); # remove trailing newlines

grep ($_=$IDIN{$_},@author); # replace authornames with ids


foreach $i (@author) {

    unless ($i == -1 || $i == 0) { # do not include non-authors
$product_byid{$i}++;
}
}
```

```
}

# $n = keys(%product_byid);
# print $n;

# CREATE ASSOCIATIVE ARRAY OF CO-AUTHORSHIPS FOR PROD > 1

open(SCIENTOMETRICS, "../data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;

for (@scientometrics) {


#SORT FIELDS

    @fields = split(/\|/);

#AUTHOR FIELDS

    @au = grep($_ =~ s/^\n*AU-\s//,@fields);
    if (@au == ()) {next;}
    $au = shift(@au);

@author = split(/;/,$au); # extract citing author names

grep (s/^\s*//,@author); # remove leading spaces
grep (s/\s*$//,@author); # remove trailing spaces
grep (s/^\n*//,@author); # remove leading newlines
grep (s/\n*$//,@author); # remove trailing newlines

grep ($_=$IDIN{$_},@author); # replace authornames with ids


# CREATE  ASSOCIATIVE ARRAY TO CAPTURE COAUTHORSHIPS

foreach $i (@author) {

    if ($product_byid{$i} <= 1) {next;}

    foreach $j (@author) {

    if ($i == $j || $product_byid{$j} <= 1) {next;}

    $coautid = $i . "-" . $j;

    $coauthor{$coautid}++;

}
}
}

# TURN ASSOCIATIVE ARRAY IN MATRIX FOR STRUCTURE
```

```
foreach $n (sort bynumber keys %product_byid) {

    @finalstring = ();

    if ($product_byid{$n} <= 1) {next;}

foreach $m (sort bynumber keys %product_byid) {

    if ($product_byid{$m} <= 1) {next;}

    $coautid = $n . "-" . $m;

    if ($coauthor{$coautid} == "") {

$coauthor{$coautid} = "0"; # purely for presentation
    }

    push(@finalstring,$coauthor{$coautid});
}

$aantal = @finalstring;

$coauthormatrix{$n} = pack("A3" x $aantal,@finalstring);

}


# PRINT FINAL CO-AUTHOR MATRIX FOR STRUCTURE

    open (STRUCTUREMATRIX,">../Co-authors_Structure_Matrix");
print STRUCTUREMATRIX ("Co-authorship matrix for Structure");
    open (STRUCTUREMATRIX,">>../Co-authors_Structure_Matrix");

foreach (sort bynumber keys %coauthormatrix) {

    print STRUCTUREMATRIX $coauthormatrix{$_};
}

$n = keys(%coauthormatrix);
# print $n;

sub bynumber { $a <=> $b; }
```

.10

```
#!/usr/bin/perl

open(ID, "<Authors.ID.Input.txt");
```

```
@ID = <ID>;

grep(s/\s+$//,@ID);

open(IDOUT, ">>Authors.ID.Input2.txt");

print IDOUT join("\n",@ID);
```

.11

```
#!/usr/bin/perl
# Count coword occurrences

# INPUT WORD FREQUENCY FILE

$/ = "\n"; # input one line per record
$\ = "\n"; # print output newline

open (IDFREQ,"<../WordID_Frequencies.txt") || die "Can’t open WordFreq: $!\n";

@idfreq = <IDFREQ>;
shift(@idfreq); # remove header line
grep(s/ˆ\n*//,@idfreq); # remove all leading newlines
grep(s/\n*$//,@idfreq); # remove all trailing newlines

for (@idfreq) {

($wordid, $rest, $wordfreq) = /(\d+)(:\s*)(\d+)/; # split string

$wordidfreq{$wordid} = $wordfreq;
}

# INPUT WORD ID FILE


    open (WORDID,"<../WordID.txt") || die "Can’t open WordID: $!\n";

    @wordid = <WORDID>;

    shift(@wordid); # remove header line
grep(s/ˆ\n*//,@wordid); # remove all leading newlines
grep(s/\n*$//,@wordid); # remove all trailing newlines

    for (@wordid) {

($pureword, $rest, $wordid) = /(\w+)(:\s*)(-*\d+)/; # split string

unless ($wordid == -1) {
$WORDID{$pureword} = $wordid; # build associative ID array
```

```
}
}


# INPUT SCIENTOMETRICS DATA FILE


$/ = ""; # enable paragraph mode
$* = 1; # enable multiline string searches
$\ = "\n"; # print output with \n

open(SCIENTOMETRICS, "../data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;



for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);

#TITLE FIELDS

    @ti = grep($_ =˜ s/ˆ\n*TI-\s//,@fields);
    if (@ti == ()) {next;}
    $ti = shift(@ti);
$ti =˜ s/-/ /g; # replace "-" by space


    @word = split(/\s+/,$ti);
grep(s/ˆ\W*\d+//,@word); # remove all leading digits
grep(s/\d+\W*$//,@word); # remove all trailing digits
grep(s/ˆ\W+//,@word); # remove leading nonwords
grep(s/\W+$//,@word); # remove trailing nonwords


# REPLACE WORDS WITH THEIR ID’S

grep($_=$WORDID{$_},@word);

# BUILD COWORD ASSOCIATIVE ARRAY ABOVE FREQUENCY TRESHOLD

foreach $i (@word) {

    if ($wordidfreq{$i} <= 10) {next;} # treshold

    foreach $j (@word) {

    if ($i == $j || $wordidfreq{$j} <= 10) {next;}

    $cowordid = $i . "-" . $j;
    $coword{$cowordid}++;
```

```
}
}
}

# CREATE MATRIX FOR STRUCTURE

foreach $n (sort bynumber keys %wordidfreq) {

    @finalstring = ();

if ($wordidfreq{$n} <= 10) {next;} # treshold

   foreach $m (sort bynumber keys %wordidfreq) {

if ($n == $m || $wordidfreq{$m} <= 10) {next;}

$cowordid = $n . "-" . $m;

     if ($coword{$cowordid} == "") {

$coword{$cowordid} = "0"; # purely for presentation
    }

    push(@finalstring,$coword{$cowordid});

}

$aantal = @finalstring;
$cowordmatrix{$n} = pack("A3" x $aantal,@finalstring);

}

# PRINT RESULTS

    open (COWORDMATRIX,">../Co-word_Structure_Matrix");
print COWORDMATRIX ("Co-word matrix for Structure");
    open (COWORDMATRIX,">>../Co-word_Structure_Matrix");

foreach (sort bynumber keys %cowordmatrix) {

    print COWORDMATRIX $cowordmatrix{$_};
}

$n = keys(%cowordmatrix);
print $n;

sub bynumber { $a <=> $b; }
```

.12

```
#!/usr/bin/perl

# PRODUCTIVITY PER INSTITUTION AND COOOPERATION BETWEEN INSTITUTIONS

# FIRST OPEN DATABASE OF UNIQUE ADDRESSES

open(IDIN, "<Institution_ID.Input.txt") || die "Can’t open:
$!\n";

@IDINPUT = <IDIN>;

for (@IDINPUT) {

($id, $uniaddress) = split(/: /,$_,2);
$uniaddress =~ s/\s*$//; # remove trailing spaces
$INSTITUTIONIDIN{$uniaddress} .= $id;
}

open(IDOUT, "<Institution_ID.Output.txt");
@IDOUTPUT = <IDOUT>;

for (@IDOUTPUT) {
($id, $uniaddress) = split(/: /,$_,2);
$uniaddress =~ s/\s*$//; # remove trailing spaces
$INSTITUTIONIDOUT{$id} .= $uniaddress;
}


# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";


# OPEN THE SCIENTOMETRIC DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;



# PROCESS EACH RECORD

for (@scientometrics) {


#SORT FIELDS

    @fields = split(/\|/);

    @cs = grep(s/^CS-\s//,@fields);
```

```
# TRANSLATE ARRAY VALUES IN STRING VALUES

    $cs = shift @cs;
    $cs =~ s/^\n\s*//; # Remove leading newline and spaces

# SPLIT CS-FIELD INTO INDIVIDUAL INSTITUTIONS

    @address = split(/;/,$cs);


INST: {    foreach $address (@address) {

($realaddress, $rest)  = split(/,|\//,$address,2);
$realaddress =~ s/^\n*\s*(\b\D.*)\s*\n*\s*(\b.*)\s*$/\1\2/;
if (!defined($realaddress)) {next;}

# LOOKUP ADDRESS IN DATABASE

&IDIN;
&IDOUT;
$counter++;
$pubaddress{$realaddress}++;

}

$coinstitution{$counter}++;
$counter = 0;

}
}

# COMPUTE DISTRIBUTION PRODUCTIVITY OVER INSTITUTIONS

foreach $realaddress (sort keys(%pubaddress)) {

    $proddist{$pubaddress{$realaddress}}++;
}

# RANK INSTITUTIONS ACCORDING TO PRODUCTIVITY

foreach $realaddress (sort keys(%pubaddress)) {

    $prodrank{$pubaddress{$realaddress}} .= $realaddress . ",";
}


# PRINT RESULTS

open(PUBINSTITUTE, ">Productivity_Institution");

print PUBINSTITUTE "Productivity per institutional address:
publications, institutions";

open(PUBINSTITUTE, ">>Productivity_Institution");
```

```
foreach $realaddress (sort keys(%pubaddress)) {

    print PUBINSTITUTE ($pubaddress{$realaddress}, " ",$realaddress);
}

open(COINSTITUTE, ">Co_Institutions");

print COINSTITUTE "Co-authorships of institutions: number of
co-institutions, number of instances";

open(COINSTITUTE, ">>Co_Institutions");

foreach $counter (sort bynumber keys(%coinstitution)) {

    print COINSTITUTE ($counter, " ", $coinstitution{$counter});
}

open(PRODDIST, ">Productivity_Distribution_Institutions");

print PRODDIST "Distribution of productivity over institutions:
productivity, number of institutions";

open(PRODDIST, ">>Productivity_Distribution_Institutions");

foreach $pubaddress (sort bynumber keys(%proddist)) {

    print PRODDIST ($pubaddress, " ", $proddist{$pubaddress});
}

open(PRODRANK, ">Ranked_Institutions");

print PRODRANK "Ranked Institutions";

open(PRODRANK, ">>Ranked_Institutions");

foreach (reverse sort bynumber keys(%prodrank)) {

    chop $prodrank{$_};
    print PRODRANK ($_, " ", $prodrank{$_}) unless ($_ < 5);
}


sub bynumber {$a <=> $b;}

sub IDIN {

$realaddress = $INSTITUTIONIDIN{$realaddress};
 if ($realaddress == 0) {next INST;}
}

sub IDOUT {

    $realaddress = $INSTITUTIONIDOUT{$realaddress};
}
```

.13

```
#!/usr/bin/perl

# CREATE DATABASE FOR ADDRESS UNIFICATION


# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;



# PROCESS EACH RECORD

for (@scientometrics) {

# ASSIGN RECORD NUMBER

#    $recno++;

#SORT FIELDS

    @fields = split(/\|/);

    @cs = grep(s/^CS-\s//,@fields);

# TRANSLATE ARRAY VALUES IN STRING VALUES

    $cs = shift @cs;
    $cs =~ s/^\n\s*//; # Remove leading newline and spaces


# SPLIT CS-FIELD INTO INDIVIDUAL INSTITUTIONS

    @address = split(/;/,$cs);


    foreach $address (@address) {
```

```
($realaddress, $rest)  = split(/,|\//,$address,2);
$realaddress =~ s/^\n*\s*(\b\D.*)\s*\n*\s*(\b.*)\s*$/\1\2/;

if (!defined($realaddress)) {next;}

if (grep(/$realaddress/,%INSTITUTIONID)) {

    next;
}
else {

$id++;
$INSTITUTIONID{$id} .= $realaddress;
}
}

}

open(ID, ">Institution_ID");
print ID "Unique ID's for each institutional address";

open(ID, ">>Institution_ID");

foreach (sort bynumber keys(%INSTITUTIONID)) {

    print ID ($_, ": ", $INSTITUTIONID{$_});
}

sub bynumber {$a <=> $b;}
```

.14

```
#!/usr/bin/perl

# Print output default newline

$\ = "\n";

# CREATE PURIFIED DATABASE OF UNIQUE ADDRESSES FOR PRINT OUTPUT

# FIRST OPEN DATABASE OF UNIQUE ADDRESSES

open(ID, "<Institution_ID.Input.txt") || die "Can't open: $!\n";

@ID = <ID>;
shift; # remove first line

for (@ID) {
```

```
    chop; # remove newline

($id, $uniaddress) = split(/: /,$_,2);
$uniaddress =˜ s/ˆ\s*//; # remove leading spaces
$INSTITUTIONID{$id} .= $uniaddress . ",";

}

# CHECK TO SEE WHETHER ADDRESS IS UNIQUE (CHECK ON COMMA)

foreach (sort keys(%INSTITUTIONID)) {

    chop $INSTITUTIONID{$_}; # remove trailing comma


if (($n = index($INSTITUTIONID{$_},",")) >= 0) {

    print $n;

    $INSTITUTIONID{$_} = substr($INSTITUTIONID{$_},0,$n);
}



# CREATE OUTPUT DATABASE

$PURIFIED{$_} .= $INSTITUTIONID{$_};

}


open(PURIFY, ">Institution_ID.Purified");
print PURIFY "Institutional ID’s, purified: ";
open(PURIFY, ">>Institution_ID.Purified");

foreach (sort bynumber keys(%PURIFIED)) {

    print PURIFY ($_, ": ", $PURIFIED{$_});
}

sub bynumber {$a <=> $b;}
```

.15

```
#!/usr/bin/perl

# RANK CITED JOURNALS PER YEAR
```

```
# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "data") || die "Can't open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;



# PROCESS EACH RECORD

for (@scientometrics) {

# ASSIGN RECORD NUMBER

    $recno++;
    $nr=0;

#SORT FIELDS

    @fields = split(/\|/);

    @py = grep(s/^PY-\s//,@fields);
    @cr = grep(s/^CR-\s//,@fields);

# TRANSLATE ARRAY VALUES IN STRING VALUES

    $py = shift @py;
    $cr = shift @cr;
    $py =~ s/^\n//; # remove leading newline
    $cr =~ s/^\n//;


# SPLIT CR-FIELD INTO INDIVIDUAL REFERENCES

    @reference = split(/^/,$cr);


   for (@reference) {


# Split individual references
# fourdigit is cited year

($begin, $citedyear, $end) = /^(.*)\b(\d{4})\b(.*)/;


next if ($citedyear < 1200 || $citedyear > $py);

# cited journal at end
```

```
($begin, $citedjournal) = /^(.*),\s+(\D+\w*)\b\d*\W*$/;

# Count number of valid references

$nr++;
$totalref++;

# MAKE ARRAY OF CITED JOURNALS

$Citedjournals{$citedjournal}++;

}
}

# Rank cited journals

foreach (sort keys(%Citedjournals)) {

    $Citedjournallist{$Citedjournals{$_}} .= $_ . ",";
}


# PRINT THE COUNTS

# Number of journals cited by Scientometrics

open(JOURNALCITED, ">Journals_cited");
print JOURNALCITED "Cited journals: journal, citations";

open(JOURNALCITED, ">>Journals_cited");

foreach $citedjournal (sort keys(%Citedjournals)) {

    $journalcitation++;

    if ($Citedjournals{$citedjournal}==1) {$journalonce++;}

    print JOURNALCITED ($citedjournal, " ", $Citedjournals{$citedjournal});

}

#  Cited journals ranked

open(JOURNALCITED2, ">Journals_cited_ranked");
print JOURNALCITED2 "Ranked cited journals: citations, journals";
open(JOURNALCITED2, ">>Journals_cited_ranked");


foreach (reverse sort bynumber keys(%Citedjournallist)) {

    chop($Citedjournallist{$_});
    print JOURNALCITED2 ($_, " ", $Citedjournallist{$_}) unless ($_ < 10);
}
```

```
sub bynumber {$a <=> $b;}
```

.16

```
#!/usr/bin/perl

# Program to introduce new field separators in DIALOG data on
# Scientometrics


# Ensure paragraph mode again

$/ = "";
$* = 1;

open (CLEANED, "<data.clean");

# Open outputfile

open(REALCLEAN, ">data.real.cleaned");
print REALCLEAN "";
open(REALCLEAN,">>data.real.cleaned");


# Replace old field separator

while (<CLEANED>) {

    s/\|*\s*\n([A-Z][A-Z]\-\s)/\|\n\1/g;
    print REALCLEAN $_;

}
```

.17

```
#!/usr/bin/perl

# Program to renew record separator Dialog DATA on Scientometrics, part II

# Enable paragraph mode

$/ = "||";
```

```
$* = 1;

# Open the database, complain if impossible

open(SCIENTOMETRICS, "data.tobecleaned") || die "Can’t open
data.tobecleaned $!\n";

# Open outputfiles, empty first

open(CLEANED, ">data.clean");
print CLEANED "";
open (CLEANED, ">>data.clean");

# Replace old paragraph separator

while(<SCIENTOMETRICS>) {

    s/(\s*\d+\/\d+\/\d+)/\n\n\1/g;
    print CLEANED $_;

}
```

.18

```
#!/usr/bin/perl
# COUNT NUMBER PUBLICATIONS PER YEAR AND PER LANGUAGE

# Enable paragraph mode, format output with newline

$/ = "";
$* = 1;
$\ = "\n";

# Open the database, complain if impossible

open(SCIENTOMETRICS, "data") || die "Can’t open data: $!\n";

$num = @scientometrics = <SCIENTOMETRICS>;

open (TOTALREC, ">total_records");
print TOTALREC $num;

for (@scientometrics) {

#Sort the fields

    @fields = split(/\|/);
```

```
    @la = grep(s/^LA-\s//,@fields);
    @py = grep(s/^PY-\s//,@fields);

# Translate array values in string values

    $la = shift @la;
    $py = shift @py;
    $py =~ s/^\n//; # remove leading newlines
    $la =~ s/^\n//;
    $py =~ s/\n$//; # remove trailing newlines


# SET UP ASSOCIATIVE ARRAYS FOR COUNTING

# Count number publications per year

     $pubyear{$py}++ unless !(defined($py));


# Count number publications per language

     $languagecount{$la}++;

}


# PRINT THE COUNTS

# Number of publications per year

open (PUBYEAR, ">Publications_year");
print PUBYEAR "";

open (PUBYEAR, ">>Publications_year");

foreach $py (sort keys(%pubyear)) {

    print PUBYEAR ($py, " ", $pubyear{$py});

}


# Number of publications per language

open (LANGUAGEPUB, ">Publications_language");
print LANGUAGEPUB "";
open (LANGUAGEPUB, ">>Publications_language");

foreach $la (sort keys(%languagecount)) {

    print LANGUAGEPUB ($la, " ", $languagecount{$la});

}
```

.19

```
#!/usr/bin/perl

# Compute purified word frequencies

# INPUT WORD FREQUENCY FILE

$\ = "\n";

open (WORDFREQ,"<../WordFrequency") || die "Can't open WordFrequency: $!\n";

@wordfreq = <WORDFREQ>;

shift(@wordfreq); # remove header line
grep(s/^\n*//,@wordfreq); # remove all leading newlines
grep(s/\n*$//,@wordfreq); # remove all trailing newlines

for (@wordfreq) {

($word, $rest, $frequency) = /(\w+)(:\s*)(-*\d+)/; # split string

$wordfreq{$word} = $frequency;
}

# INPUT WORD ID FILE


    open (WORDID,"<../WordID.txt") || die "Can't open WordID: $!\n";

    @wordid = <WORDID>;

    shift(@wordid); # remove header line
grep(s/^\n*//,@wordid); # remove all leading newlines
grep(s/\n*$//,@wordid); # remove all trailing newlines

    for (@wordid) {

($pureword, $rest, $wordid) = /(\w+)(:\s*)(-*\d+)/; # split string

unless ($wordid == -1) {
$WORDID{$pureword} = $wordid; # build associative ID array

}
}

# COMBINE BOTH ARRAYS

foreach $i (sort keys %WORDID) {
```

```
    $realfreq{$WORDID{$i}} = $realfreq{$WORDID{$i}} + $wordfreq{$i};


}

# PRINT RESULTS

open (IDFREQ,">../WordID_Frequencies");
print IDFREQ "WordID : Frequencies";

open (IDFREQ,">>../WordID_Frequencies");

foreach (sort bynumber  keys %realfreq) {

    print IDFREQ ($_, ": ", $realfreq{$_});
}

sub bynumber { $a <=> $b; }
```

.20

```
#!/usr/bin/perl

# SORT NUMBER OF REFERENCES AND COMPUTE PRICE INDEX FOR SCIENTOMETRICS


# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "data") || die "Can’t open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;



# PROCESS EACH RECORD

for (@scientometrics) {

# ASSIGN RECORD NUMBER

    $recno++;
    $nr=0;
```

```
#SORT FIELDS

    @fields = split(/\|/);

    @py = grep(s/^PY-\s//,@fields);
    @cr = grep(s/^CR-\s//,@fields);

# TRANSLATE ARRAY VALUES IN STRING VALUES

    $py = shift @py;
    $cr = shift @cr;
    $py =~ s/^\n//; # remove leading newline
    $cr =~ s/^\n//;


# SPLIT CR-FIELD INTO INDIVIDUAL REFERENCES

    @reference = split(/^/,$cr);


   for (@reference) {


# Split individual references
# fourdigit is cited year

($begin, $citedyear, $end) = /^(.*)\b(\d{4})\b(.*)/;


next if ($citedyear < 1200 || $citedyear > $py);

# cited journal at end

($begin, $citedjournal) = /^(.*),\s+(\D+\w*)\b\d*\W*$/;

# Count number of valid references

$nr++;
$totalref++;

# MAKE ARRAY OF CITED JOURNALS

$Citedjournals{$citedjournal}++;

# COMPUTE PRICE INDEX ACCORDING TO MOED PER CITING ARTICLE

$difference = $py - $citedyear;

if (($difference < 6) && ($difference >= 0)) {$moed1++;}

# Make array of relative ages

$relativeage{$difference}++;

# COMPUTE PRICE INDEX ACCORDING TO PRICE
```

```
 $priceindexprice{$py} .= $difference . " ";

}

# MAKE ARRAY OF NUMBER REFERENCES

    $refnumber{$nr}++;


# COMPUTE PRICE INDEX ACCORDING TO MOED PER CITING ARTICLE


if ($nr) {

    $nr == $moed1 ? $priceindart = 100 : $priceindart = ($moed1/$nr * 100);

}

else {$priceindart = -1;}


$priceindexmoed = $priceindexmoed + $priceindart;

$priceindart = int($priceindart);

$distpricemoed{$priceindart}++;

$moed1 = 0;

}

# NOW COMPUTE OVER ALL RECORDS

$PRICEM = $priceindexmoed/@scientometrics;

# Compute Price Index according to Price (for each year)

 foreach $py (sort keys(%priceindexprice)) {

     @differences = split(/\s/,$priceindexprice{$py});
$counter1=0;
$counter2=0;

foreach $difference (@differences) {

    $difference =~ s/\D//g;

    $counter1++;

    if ($difference < 6) {$counter2++;}
}

$yearindex = int($counter2/$counter1 * 100);
$yearreferences = $counter1;
```

```
    $PRICEP{$py} .= $yearindex . " ";

}

foreach (sort keys(%PRICEP)) {

    $totalprice = $totalprice + $PRICEP{$_};
$counter3++;

}

$PRICEPRICE = $totalprice/$counter3;


# Rank cited journals

foreach (sort keys(%Citedjournals)) {

    $Citedjournallist{$Citedjournals{$_}} .= $_ . ",";
}


# PRINT THE COUNTS

open (PRICEINDEXM, ">PriceIndexMoed");
print PRICEINDEXM ("Overall Price Index according to Moed: ",$PRICEM);

open(PRICEINDEXP, ">PriceIndexPrice");
print PRICEINDEXP ("Overall Price Index according to Price: ",$PRICEPRICE);

# Price Index per publication year

open(INDEXYEAR, ">PriceIndex_year");
print INDEXYEAR "Price Index per publication year: year, Price Index";
open(INDEXYEAR, ">>PriceIndex_year");

foreach $py (sort keys(%PRICEP)) {

    print INDEXYEAR ($py, " ", $PRICEP{$py});
}

# Distribution Moed’s Price Index over articles

open(DISTMOED, ">Distribution_Moed");

print DISTMOED "Distribution of Price Index (Moed) over citing
articles: Price Index, articles";

open(DISTMOED, ">>Distribution_Moed");

foreach (sort bynumber keys(%distpricemoed)) {

    print DISTMOED ($_, " ", $distpricemoed{$_});
}
```

```
# Distribution number of references over articles

open(REFNUMBER, ">References_article");
print REFNUMBER "Distribution of number of references over articles: references, artic

open(REFNUMBER, ">>References_article");

foreach $nr (sort bynumber keys(%refnumber)) {

    print REFNUMBER ($nr, " ", $refnumber{$nr});

}

# Number of journals cited by Scientometrics

open(JOURNALCITED, ">Journals_cited");
print JOURNALCITED "Cited journals: journal, citations";

open(JOURNALCITED, ">>Journals_cited");

foreach $citedjournal (sort keys(%Citedjournals)) {

    $journalcitation++;

    if ($Citedjournals{$citedjournal}==1) {$journalonce++;}

    print JOURNALCITED ($citedjournal, " ", $Citedjournals{$citedjournal});

}

#  Cited journals ranked

open(JOURNALCITED2, ">Journals_cited_ranked");
print JOURNALCITED2 "Ranked cited journals: citations, journals";
open(JOURNALCITED2, ">>Journals_cited_ranked");


foreach (reverse sort bynumber keys(%Citedjournallist)) {

    chop($Citedjournallist{$_});
    print JOURNALCITED2 ($_, " ", $Citedjournallist{$_}) unless ($_ < 10);
}

open(TOTALREF, ">Total_references");
print TOTALREF ("Total number of valid references: ",$totalref);

open(JOURNALCIT, ">Total_journals_cited");

print JOURNALCIT ("Total number of unique cited journals: ",
$journalcitation, "\nTotal number of journals cited only once: ",
$journalonce);

open(RELATIVEAGE, ">Distribution_age_cited");
```

```
print RELATIVEAGE "Distribution of relative age of cited articles:
number of articles, age";
open(RELATIVEAGE, ">>Distribution_age_cited");

foreach (sort bynumber keys(%relativeage)) {

    print RELATIVEAGE ($_, " ", $relativeage{$_});
}

sub bynumber {$a <=> $b;}
```

.21

```
#!/usr/bin/perl
# Count title word frequencies


# ENABLE PARAGRAPH MODE, FORMAT OUTPUT WITH NEWLINE

$/ = "";
$* = 1;
$\ = "\n";

# OPEN THE DATABASE, COMPLAIN IF IMPOSSIBLE

open(SCIENTOMETRICS, "../data") || die "Can't open data: $!\n";

@scientometrics = <SCIENTOMETRICS>;



for (@scientometrics) {

#SORT FIELDS

    @fields = split(/\|/);

#TITLE FIELDS

    @ti = grep($_ =~ s/^\n*TI-\s//,@fields);
    if (@ti == ()) {next;}
    $ti = shift(@ti);
$ti =~ s/-/ /g; # replace "-" by space


    @word = split(/\s+/,$ti);
grep(s/^\W*\d+//,@word); # remove all leading digits
grep(s/\d+\W*$//,@word); # remove all trailing digits
```

```
grep(s/^\W+//,@word); # remove leading nonwords
grep(s/\W+$//,@word); # remove trailing nonwords


    foreach $word (@word) {


if (defined($word) && $word ne "") {

$frequency{$word}++; # count word frequency
    }
    }
}

# PRINT FREQUENCY ARRAY

open (FREQUENCY, ">../WordFrequency");
print FREQUENCY "Word frequencies in titles";
open (FREQUENCY, ">>../WordFrequency");

foreach (sort keys %frequency) {

    print FREQUENCY ($_, ": ", $frequency{$_});

}

# INVERT FREQENCY ARRAY

foreach (sort keys %frequency) {

    $freqinvert{$frequency{$_}} .= $_ . " ";

}


# PRINT INVERTED FREQUENCY ARRAY

open (FREQUENCY, ">../WordFrequencyInverted");
print FREQUENCY "Word frequencies in titles";
open (FREQUENCY, ">>../WordFrequencyInverted");

foreach (sort bynumber keys %freqinvert) {

    print FREQUENCY ($_, ": ", $freqinvert{$_});

}

# CREATE ID ARRAY TO USE AS FILTER

    foreach (sort keys %frequency) {

$counter++;
$wordid{$_} = $counter;
    }
```

```
# PRINT WORDID ARRAY

open (WORDID, ">../WordID");
print WORDID "Title Word ID’s; Word : ID";
open (WORDID, ">>../WordID");

foreach (sort keys %wordid) {

    print WORDID ($_, ": ", $wordid{$_});

}

# SUBS

sub bynumber { $a <=> $b;}
```

Bibliography

A to F G to L M To R S To Z

Aaronson, S. (1975), 'The footnotes of science', *Mosaic* **6**(2), 22–27.

Abelson, P. H. (1966), 'Coping with the information explosion', *Science* **154**(3745), xxx.

Adair, W. C. (1954), Citation indexes for scientific literature. Second draft, 6 pp.

Adair, W. C. (1955), 'Citation indexes for scientific literature?', *American Documentation* **6**, 31–32.

Ankersmit, F. R. (1990), *De navel van de geschiedenis. Over interpretatie, representatie en historische realiteit*, Historische uitgeverij Groningen.

Anonymous (1956), 'Needed—a documentation center. organizing the voluminous scientific information in the u.s. can produce new and vital data', *Chemical & Engineering News* pp. 514–516.

Anonymous (1959 or 1960), 'Reviewer's comments on garfield's citation index'. Personal Archive Eugene Garfield, 5 pages.

Anonymous (1964*a*), 'From the reviewers: Genetics citation index', *Science Fortnightly* **1**(14), 4.

Anonymous (1964*b*), 'Recent publications: Genetics citation index', *Eugenetics Quarterly* **11**(3), 185.

Anonymous (1966), 'Drop more names', *Nature* **211**(5049), 556–557.

Anonymous (1983*a*), "Hef medische faculteit VU op", *NRC/Handelsblad* .

Anonymous (1983*b*), 'Medische faculteit in Rotterdam aan de top bij medisch onderzoek', *NRC/Handelsblad* .

Anonymous (1984), 'Commentaar van het NcGv op het Advies inzake de Prioriteiten in het Gezondheidsonderzoek van de Raad van Advies voor het Wetenschapsbeleid van augustus en de twee achtergrondstudies', Letter to the Ministry of WVC. 02314/TVDG/JL.

Anonymous (1985), 'Onderzoek naar kanker en hartziekten moet geneeskunde optimaliseren. kabinet grotendeels eens met RAWB-advies 'prioriteiten gezondheidszorg'', *De Nederlandse Staatscourant* (107).

ANP (1983), 'Rotterdamse universiteit blinkt uit met medische onderzoeken', *Algemeen Dagblad* .

Appleby, J., Hunt, L. & Jacob, M. (1994), *Telling the truth about history*, Norton, New York.

Ashmore, M. (1989), *The reflexive thesis: Wrighting sociology of scientific knowledge.*, The University of Chicago Press.

Bakker, P. & Rigter, H. (n.d.), 'Editors of medical journals: who and from where?'.

Barinova, Z. et al. (1967), 'Izucheniye nauchnykh zhurnalov kak kanalov svyazi. otsenka vlada, vnosymogo otdelnymi stranami v mirovoy nauchny potok (studying journals as channels of communication. evaluation of the contributions of particular countries to the world information circuit)', *Nauchno-Tekhnicheskaya Informatsiya* **2**(12), 1–11.

Barrett, R. L. & Barrett, M. A. (1957), 'Journals most cited by chemists and chemical engineers', *Journal of Chemical Education* **34**(1), 35–38.

Bateson, G. (1980), *Mind and Nature*, Bantam, New York.

Bazerman, C. (1988), *Shaping written knowledge: The genre and activity of the experimental article in science.*, University of Wisconsin Press, Madison.

Beardmore, J. A. (1964), 'Genetics citation index', *Genetica* **35**(4), 378–379.

Beck, M. T. (1978), 'Editorial statements', *Scientometrics* **1**(1), 3–4.

Bensing, D. J. M. (1984). Letter to the RAWB.

Bernal, J. D. (1939), *The Social Function of Science*, Routledge & Kegan Paul Ltd., London.

Bernal, J. D. (1948), The information problem in science, *in* 'Royal Society Scientific Information Conference', Royal Society London.

Bernal, J. D. (1964), After twenty-five years, *in* A. M. Maurice Goldsmith, ed., 'The Science of Science', Souvenir Press Ltd.

Bernal, J. D. (1965), 'Science citation index', *Science Progress* **53**(211), 455–459.

Biochemie, V. (1982), *Over leven. Betekenis van de biochemie in Nederland*, Staatsuitgeverij, 's Gravenhage, the Netherlands.

Blom, T. (1997), *Complexiteit en Contingentie. Een kritische inleiding tot de sociologie van Niklas Luhmann*, Kok/Agora, Kampen.

Bloor, D. (1976), *Knowledge and Social Imagery*, Chicago University Press.

Blume, S. (1974), *Towards a political sociology of science*, Free Press, New York.

Blume, S. (1986), *The development of Dutch science policy in international perspective, 1965–1985*, Serie Achtergrondstudies, nr. 14, Ministerie van Onderwijs en Wetenschappen, The Hague.

Board, N. S. (1973), Science indicators 1972, Technical report, NSB, Washington DC.

Boel, J. (1978), Beoordeling van (universitair) wetenschappelijk onderzoek: een probleemschets en een wijze van aanpak. CAVWO/ZWO-symposium, Jaarbeurshallen, Utrecht.

Brand, S. (1987), *The Media Lab. Inventing the future at M.I.T.*, Penguin Books, New York.

Britton, J. P. (1964), The productivity of scientists: A prelude to manpower studies. Derek de Solla Price Collection, Paris.

Broadus, R. N. (1967), 'Early approaches to bibliometrics', *Journal of the American Society for Information Science* **38**(2), 127–129.

Brodman, E. (1944), 'Choosing physiology journals', *Medical Library Association Bulletin* **XXXII**, 479–483.

Brookes, B. C. (1988), Biblio-, sciento-, inform-metrics??? what are we talking abaout?, *in* 'Informetrics 89/90', Elsevier, Amsterdam, pp. 31–43.

Bukharin, N. et al. (1931), *Science at the Cross Roads*, Russian Foreign-Languages Press, Kniga. Reprinted 1971 by Frank Cass & Co. Ltd.

Burt, R. S. (1982), *Toward a Structural Theory of Action.*, Academic Press, New York.

Burt, R. S. & Minor, M. J. (1983), *Applied Network Analysis*, Sage, Beverly Hills.

Burton, R. E. (1959*a*), 'Citations in American Engineering Journals. part i', *American Documentation* **10**(1), 70–73.

Burton, R. E. (1959*b*), 'Citations in American Engineering Journals. part ii', *American Documentation* **10**(2), 135–137.

Bush, V. (1945), *Science: The Endless Frontier*, US Government Printing Office, Washington DC. Charter Document for the US National Science Foundation.

Callon, M. (1986), Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieux Bay, *in* J. Law, ed., 'Power, Action and Belief: A new sociology of knowledge?', Sociological Review Monograph, Routledge & Kegan Paul, London, pp. 196–229.

Callon, M., Courtial, J.-P., Turner, W. & Bauin, S. (1983), 'From translations to problematic networks: An introduction to co-word analysis', *Social Science Information* **22**, 191–235.

Cason, H. & Lubotsky, M. (1936), 'The influence and dependence of psychological journals on each other', *Psychological bulletin* **33**, 95–103.

Chang, H. (1975), Evaluation and survey of a subfield of physics. magnetic resonance and relaxation studies in the netherlands, Technical report, FOM, Utrecht, The Netherlands.

Chang, H. & Dieks, D. (1976), 'The Dutch output of publications in physics', *Research Policy* **5**, 380–396.

Charlton, L. (1981), 'An index to getting ahead', *The New York Times* .

Chubin, D. E. & Moitra, S. (1975), 'Content analysis of references: Adjunct or alternative to citation counting', *Social Studies of Science* **5**, 423–441.

Chubin, D. & Hackett, E. J. (1990), *Peerless Science: Peer review and U.S. science policy*, State University of New York Press, Albany, US.

Cleverdon, C. W. (1964), 'Citation indexing', *Nature* **203**(4944), 446–447.

Cleverdon, C. W. (1965), 'Citation indexing', *Nature* **208**(5012), 717.

Cole, F. J. & Eales, N. B. (1917), 'The history of comparative anatomy', *Science Progress* **XI**, 578–596.

Cole, J. (1970), 'Patterns of intellectual influence in scientific research', *Sociology of Education* **43**, 377–403.

Cole, J. & Cole, S. (1971), 'Measuring the quality of sociological research: Problems in the use of the *Science Citation Index*', *The American Sociologist* pp. 23–29.

Cole, R. C. (1952), 'Periodical literature for electrical engineers', *The Journal of Documentation* **8**(4), 209–226.

Cole, S. & Cole, J. (1967), 'Scientific output and recognition', *American Sociological Review* **32**, 377–390.

Collins, H. (1991), 'Captives and victims: Comments on Scott, Richards, and Martin', *Science, Technology & Human Values* **16**(2), 249–251.

Cozzens, S. (1981), 'Taking the measure of science: A review of citation theories', *International Society for the Sociology of Knowledge Newsletter* **7**(1–2), 16–20.

Cozzens, S. (1985), 'Comparing the sciences: Citation context analysis of papers from neuropharmacology and the sociology of science', *Social Studies of Science* **15**, 127–153.

Cozzens, S. (1989), 'What do citations count? the rhetoric-first model', *Scientometrics* **15**, 437–447.

Cozzens, S. E., Healy, P., Rip, A. & Ziman, J. (1990), *The research system in transition*, Kluwer, Dordrecht.

Crane, D. (1972), *Invisible Colleges*, University of Chicago Press, Chicago.

Cronin, B. (1981), 'The need for a theory of citing', *Journal of Documentation* **37**(1), 16–24.

Cronin, B. (1984), *The Citation Process. The role and significance of citations in scientific communication*, Taylor Graham, London.

Davies, M. (1966), 'Drop more names', *Nature* **211**(5052), 903. Harvard Medical School, Boston, Mass.

de Candolle, A. (1885), *Histoire des Sciences et des Savants*, 2 edn, H. Georg, Geneva/Basel.

de Ruiter, L. & Vink, H. J. (1978), 'Uitnodiging deelname CAVWO/ZWO-symposium Beoordeling wetenschappelijk onderzoek', Voorburg. Kenmerk nr. 700 bis.

de Solla Price, D. (1969), 'Ethics of citations', *Special Libraries* p. 468.

de Solla Price, D. & Gürsey, S. (1976*a*), 'Studies in scientometrics, part 1: Transience and continuance in scientific authorship', *International Forum on Information and Documentation* **1**(2), 17–24. International Federation for Documentation, Moscow.

de Solla Price, D. & Gürsey, S. (1976*b*), 'Studies in scientometrics, part 2: The relation between source author and cited author populations', *International Forum on Information and Documentation* **1**(3), 19–22. International Federation for Documentation, Moscow.

de Wilde, R. (1992), *Discipline en legende. De identiteit van de sociologie in Duitsland en de Verenigde Staten 1870–1930*, Van Gennep, Amsterdam.

Dickson, D. (1984), *The new politics of science*, Pantheon Books, New York.

Diepenhorst (1966), 'Installatierede voor de Raad voor het Wetenschapsbeleid'. DGW 149411/HOW Archief nr. 129.

Dippel, D. H. W. (1983), 'Sectie tandheelkunde academische raad/interfacultair overleg tandheelkunde', Letter to the RAWB. SU 8249.

Dits, H. (1987*a*), 'Strategie m.b.t. indicatoren'. Internal memo.

Dits, H. (1987*b*), 'Voorstel voor Actieplan Werkgroep Indicatoren'. Internal memo.

Dobrov, G. M. (1964), 'O predvidenii razvitiya nauki (on the foreseeing of the development of science)', *Vopsosv Filosofii* **10**, 71–82.

Dobrov, G. M. (1966), *Nauka o Nauke*, Kiev, Naukova Dumka.

Dyson, G. M. (1952), 'The preservation and availability of chemical knowledge', *Journal of Chemical Education* **29**, 239–243.

Earle, P. & Vickery, B. (1969), 'Social science literature use in the uk as indicated by citations', *Journal of Documentation* **25**(2), 123–141.

Edge, D. (1977), 'Why I am not a co-citationist', *4S Newsletter* **2**(3), 13–19.

Edge, D. (1979), 'Quantitative measures of science: A critical review', *History of Science* **17**, 102–134.

Edge, D. O. & Mulkay, M. J. (1976), *Astronomy Transformed: The Emergence of Radio Astronomy in Britain*, Wiley, New York.

Egghe, L. & Rousseau, R. (1990), *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*, Elsevier, Amsterdam.

Elkana, Y., Lederberg, J., Merton, R. K., Thackray, A. & Zuckerman, H. (1979), *Toward a Metric of Science: The Advent of Science Indicators*, John Wiley & Sons, New York.

England, J. M. (1982), *A Patron for Pure Science. The National Science Foundation's Formative Years 1945–57*, National Science Foundation, Washington D.C.

Franklin, J. J. (1988), 'Testing and using quantitative methods in science policy contexts: A response to Hicks', *Social Studies of Science* **18**, 365–375.

Fussler, H. H. (1949), 'Characteristics of the research literature used by chemists and physicists in the united states', *The Library Quarterly* **19**, 19–35, 119–143.

Galjaard, H. (1983), '" ... stimuleren tot nadenken ... "', *Medisch Contact* **38**(43), 1377–1378.

Galton, F. (1874), *English Men of Science*, London.

Garfield, E. (1954*a*), Association-Of-Ideas Techniques in Documentation. Shepardizing the Scientific Literature. 11 pp.

Back to top of Bibliography

Garfield, E. (1954*b*), Shepardizing the Scientific Literature. Unpublished paper, Columbia University Library School.

Garfield, E. (1955), 'Citation Indexes for Science. A New Dimension in Documentation through Association of Ideas', *Science* **122**, 108–111.

Garfield, E. (1956*a*), 'Citation indexes—new paths to scientific knowledge', *Chemical Bulletin* **43**(4), 11–12.

Garfield, E. (1956*b*), Citation indexes to the old testament. 5 pp.

Garfield, E. (1956*c*), The organization of scientific information for retrieval and dissemination. Proposal submitted to NSF.

Garfield, E. (1957), 'Breaking the subject index barrier—a citation index for chemical patents', *Journal of Patent Office Society* **XXXIX**, 583–595.

Garfield, E. (1958*a*), A general feasibility study of citation indexes for science. a proposal for research. Prepared by Eugene Garfield Associates. 1523 Spring Garden Street, Philadelphia 30, Pennsylvania. Rough draft. 6 pp.

Garfield, E. (1958*b*), A general feasibility study of citation indexes for science. a proposal for research. Prepared by Eugene Garfield Associates. 1523 Spring Garden Street, Philadelphia 30, Pennsylvania. 11 pp.

Garfield, E. (1959), A unified index to science, *in* 'Proceedings of the International Conference on Scientific Information, Washinton D.C., November 16–21, 1958', Vol. 1, National Academy of Sciences—National Research Council, Washington D.C., pp. 461–474.

Garfield, E. (1960*a*), Citation Index to Genetics and General Science Literature. Research Proposal by the Institute for Scientific Information. Submitted July 15, 1960 to the National Science Foundation.

Garfield, E. (1960*b*), Citation index to genetics literature. Application for Research Grant. Personal Archive Eugene Garfield, Philadelphia, USA.

Garfield, E. (1961), Citation index project. progress report # 1. Personal Archive Eugene Garfield, Philadelphia, USA.

Garfield, E. (1962*a*), Application of newspaper publishing methods to the dissemination of scientific information. Proposal by the Institute for Scientific Information, 33 South 17th St., Philadelphia, Pa.

Garfield, E. (1962*b*), Citation index project. progress report # 2. Personal Archive Eugene Garfield, Philadelphia, USA.

Garfield, E. (1963), *Science Citation Index. An International Interdisciplinary Index to the Literature of Science*, vol. 1: aabe-capl, 1961 edn, Institute for Scientific Information, Philadelphia 3, Pa. Pp. xxx+496. Five volume set: $ 700.

Garfield, E. (1964), "'science citation index"— A New Dimension in Indexing', *Science* **144**, 649–654.

Garfield, E. (1965), *Science Citation Index 1964 Annual Cumulation. An International Interdisciplinary Index to the Literature of Science and Technology in Eight Parts*, Institute for Scientific Information, Philadelphia. $ 1,250.

Garfield, E. (1970), 'Citation indexing for studying science', *Nature* **227**, 669–671.

Garfield, E. (1979), *Citation Indexing*, ISI, Philadelphia.

Garfield, E. & Hayne, R. L. (1955), Needed—a national science intelligence & documentation center. Presented before the AAAS, Atlanta, Ga., December, 1955.

Garfield, E. & Sher, I. H. (1963), *Genetics Citation Index—Experimental Citation Indexes to Genetics with Special Emphasis on Human Genetics*, Prepared by the Institute for Scientific Information, Philadelphia 3, Pa. Pp. XIX, 864, $ 100,- paper bound.

Garfield, E. & Sher, I. H. (1966), New tools for improving and evaluating the effectiveness of research, *in* M. C. Yovits, D. M. Gilford, R. H. Wilcox, E. Staveley & H. D. Lemer, eds, 'Research Program Effectiveness. Proceedings of the Conference Sponsored by the Office of Naval Research Washington D.C., July 27-29, 1965', Gordon and Breach, New York, pp. 135–146.

Garfield, E. et al., eds (1984), *ISI Atlas of Science. Biotechnology and Molecular Genetics 1981/82 covering 127 Research Front Specialties including 1983/84 Supplements*, ISI, Philadelphia, Pennsylvania, USA.

Garvey, W. D. & Griffith, B. C. (1964), 'The structure, objectives, and findings of a study of scientific information exchange in psychology', *American Documentation* pp. 258–267.

Gieryn, T. F. (1983), 'Boundary work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists', *American Sociological Review* **48**, 781–795.

Gilbert, N. (1976), Measuring science - some issues and problems with regard to growth indicators. a contribution to the international symposium on quantitative methods in the history of science, berkeley california, august 25-27, 1976. As discussed on 25 November 1976 by SISWO-GWO.

Gilbert, N. G. (1977), 'Referencing as persuasion', *Social Studies of Science* **7**, 113–122.

Gilbert, N. G. & Woolgar, S. (1974), 'The quantitative study of science: An examination of the literature', *Science Studies* **4**(3), 279–294.

Gilyarevsky, R. S., Multchenko, Z. M., Terekhin, A. T. & Cherny, A. I. (1968), Opyt isucheniya "science citation index" (the experience of studying "science citation index"), *in* V. V. Nalimov, ed., 'Prikladnaya Dokumentalistika', Moskva.

Golden, W. T. (1988), *Science and Technology Advice to the President, Congress and Judiciary*, Pergamon Press, New York.

Goldsmith, M. (1965), 'The science of science foundation', *Nature* **205**, 10.

Goldsmith, M. & Mackay, A. (1964), *The Science of Science*, Penguin Books, Middlesex, England.

Goudsblom, J. (1962,1970), *Nihilisme en Cultuur*, Atheneum-Polak en Van Gennep, Amsterdam.

Graham, L. R. (1993), *Science in Russia and the Soviet Union. A short history*, Cambridge University Press.

Greenberg, D. S. (1967), *The Politics of Pure Science. An Inquiry into the Relationship between Science & Government in the United States*, New American Library, New York.

Gross, P. L. K. & Gross, E. M. (1927), 'College libraries and chemical education', *Science* **LXVI**(1713), 385–389.

Gross, P. L. K. & Woodford, A. C. (1931), 'Serial literature used by american geologists', *Science* **LXXIII**, 660–664.

Gurjeva, L. G. (1992), The Early Soviet Scientometrics and Soviet Scientometricians, Master's thesis, Science Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam.

Guzzetti, L. (1995), *A Brief History of European Union Research Policy*, European Commission, Brussels.

Haanen, C. (1983), Letter to Dr. H. G. van Bueren, RAWB, The Hague. Kliniek voor Inwendige Ziekten, afdeling Bloedziekten. SintRadboudziekenhuis, Katholieke Universiteit Nijmegen.

Hacking, I. (1983), *Representing and intervening: Introductory topics in the philosophy of natural science*, Cambridge University Press, Cambridge.

Hagendijk, R. (1996), Wetenschap, Constructivisme en Cultuur, PhD thesis, University of Amsterdam, Amsterdam.

Haraway, D. J. (1991), *Simians, Cyborgs, and Women—The Reinvention of Nature*, Free Associations Books, London.

Harnad, S. (1990), 'Scholarly skywriting and the prepublication continuum of scientific inquiry', *Psychological Science* **1**, 342–343. ftp://cogsci.ecs.soton.ac.uk/pub/harnad/Harnad/harnad90.skywriting.

Harnad, S. (1991), 'Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge', *The Public-Access Computer Systems Review* **2**(1), 39–53. http://info.lib.uh.edu/pacsrev.html.

Hart, H. C. (1949), 'Re: Citation system for patent office', *Journal of the Patent Office Society* **31**, 714.

Hearings, House of Representatives, Committee on Appropriations (1957).

Heinisch, O. (1965), 'Garfield, E., und I.H. Sher (Herausgeber): Genetics Citation Index', *Biometrischen Zeitschrift* **7**(2), 125.

Hicks, D. (1987), 'Limitations of co-citation analysis as a tool for science policy', *Social Studies of Science* **17**, 295–316.

Hicks, D. (1988), 'Limitations and more limitations of co-citation analysis/bibliometric modelling: A reply to Franklin', *Social Studies of Science* **18**, 375–384.

Hicks, D. & Potter, J. (1991), 'Sociology of scientific knowledge: A reflexive analysis of science disciplines and disciplining science', *Social Studies of Science* **21**, 459–501.

Hofstadter, R. (1962), *Anti-intellectualism in American life*, Vintage Books, New York.

HOW (1966), 'Tweede bijeenkomst interdepartementaal overleg voor het wetenschapsbeleid. bijlage 3: verslag eerste bijeenkomst'. Inventaris HOW, nr. 661.

Hoyningen-Huene, P. (1993), *Reconstructing scientific revolutions*, The University of Chicago Press.

HSTC (1980), *Toward the Endless Frontier. History of the Committee on Science and Technology, 1959–79*, US Government Printing Office, Washington DC.

Huizinga, J. (1937), *De wetenschap der geschiedenis*, Tjeenk Willink, Haarlem.

Hutter, W. (1983*a*), 'Commentaar op enkele opmerkingen van kritiek n.a.v. het RAWB-advies 'Prioriteiten in het gezondheidsonderzoek''. RW-1627.

Hutter, W. (1983*b*), 'Conceptadvies inzake rapport VC Biochemie'. RW 1552, Raadsvergadering 6-5-1983, Agendapunt 6.

Hutter, W. & Rigter, H. (1983), 'Ten geleide bij de achtergrondstudies i en ii van het projekt Prioriteiten in het gezondheidsonderzoek (rw 1529 en 1536)'.

Institute for Scientific Information (1964), 'What is the *Science Citation Index*?', Philadelphia.

ISI (1961), 'The citation index', *The Journal of Heredity* **52**(4), 182.

Jasanoff, S. (1990), *The fifth branch: Science advisers as policymakers*, Harvard University Press, Cambridge, MA.

Jolles, J. (1978), Beoordeling van (universitair) wetenschappelijk onderzoek: een probleemschets en een wijze van aanpak. CAVWO/ZWO-symposium, Jaarbeurshallen, Utrecht.

Kaplan, N. (1965), 'Prolegomena to the footnote', *American Documentation* **16**, 179–187.

Katz, J. S. (1995), *Desktop Scientometrics*, in *STEEP Special Report No 3* (Katz et al. 1995), chapter Appendix A, pp. 87–104.

Katz, J. S. et al. (1995), *The Changing Shape of British Science*, STEEP Special Report No 3, SPRU, University of Sussex.

Kersten, A. E. (1996), *Een organisatie van en voor onderzoekers. De Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (Z.W.O.) 1947-1988*, Van Gorcum, Assen.

Kessler, M. M. (1961), An experimental study of bibliographic coupling between technical papers. Preprint.

Kessler, M. M. (1965), 'The MIT Technical Information Project', *Physics Today* pp. 28–36.

Kessler, M. M. & Heart, F. E. (1962), Concerning the probability that a given paper will be cited. Supported by a grant from NSF.

Knorr-Cetina, K. (1981), *The manufacture of knowledge: An essay on the constructivist and contextual nature of science.*, Pergamon, Oxford.

Koefoed, P. A. (1976), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 10. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koefoed, P. A. (1978*a*), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 32/26. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koefoed, P. A. (1978*b*), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 33. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koefoed, P. A. (1978*c*), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 34/27. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koefoed, P. A. (1978*d*), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 42/31. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koefoed, P. A. (1978*e*), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 46. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koefoed, P. A. (1979), SISWO groep wetenschapsonderzoekers nieuwsbrief nr. 56. SISWO, OZ Achterburgwal 128, Amsterdam. Copies from Arie Rip's Personal Archive.

Koeze, P. (1974), De Nederlandse natuurkundigen. enige statistische gegevens over de afstudeer-generatie 1920 tot 1971, Technical report, NNV/FOM, Utrecht, The Netherlands.

Krauze, T. et al. (1977), The sociology of science in Poland, *in* Merton & Gaston (1977), pp. 193–223.

Langendorff, T. (1986), Informatievoorziening ten behoeve van het wetenschapsbeleid. Internal memo.

Larkey, S. V. (1949), 'The Army Medical Library Research Project at the Welch Medical Library', *Bulletin of the Medical Library Association* **37**, 121–124.

Latour, B. (1984), *Les Microbes: Guerre et Paix suivi de Irréductions*, Editions A. M. Métaillié.

Latour, B. (1987), *Science in action: How to follow scientists and engineers through society.*, Harvard University Press, Cambridge, MA.

Latour, B. (1988), The politics of explanation: an alternative, *in* Woolgar (1988*a*), chapter 8, pp. 155–177.

Latour, B. & Woolgar, S. (1986), *Laboratory Life: The Construction of Scientific Facts*, 2nd edn, Princeton University Press.

le Pair, C. (1969), 'Coördinatie van het wetenschapsbeleid in de natuurkunde', *Universiteit en Hogeschool* **16**(2), 113–123.

le Pair, C. (1970), Lezing voor NSF. Draft notices.

le Pair, C. (1974), 'Wetenschapsbeleid met betrekking tot de natuurkunde', Studium Generale Technische Hogeschool Eindhoven. Dictaatnummer 9.011.

le Pair, C. (1975), 'Peer review, fair of unfair?', *Universiteit en Hogeschool* **22**(3), 177–182.

le Pair, C. (1976*a*), 'Conferentie over quantitative methods in the history of science, berkeley, aug. 1976. bezoek aan enkele instituten in amerika', FOM-41488.

le Pair, C. (1976*b*), 'De vergelijkende beoordeling van onderzoek', *FOM Publicaties* pp. 15–25. FOM - 40345/1.

Lederberg, J. (1962*a*), Notes on a technical information system. Draft, Personal Archive Eugene Garfield, Philadelphia, USA.

Lederberg, J. (1962*b*), Scitel: A central scientific communication system. Draft, Personal Archive Eugene Garfield, Philadelphia, USA.

Lemaine, G., MacLeod, R., Mulkay, M. & Weingart, P. (1976), *Perspectives on the Emergence of Scientific Disciplines*, Mouton-Aldine, The Hague.

Leydesdorff, L. (1986), 'Syllabus scientometrie', Department of Science Dynamics, Amsterdam.

Leydesdorff, L. (1987), 'Towards a theory of citation', *Scientometrics* **12**, 287–291.

Leydesdorff, L. (1989), 'Words and co-words as indicators of intellectual organization', *Research Policy* **19**, 209–223.

Leydesdorff, L. (1995), *The Challenge of Scientometrics*, DSWO Press, Leiden.

Leydesdorff, L. & Amsterdamska, O. (1990), 'Dimensions of citation analysis', *Science, Technology & Human Values* **15**, 305–335.

Leydesdorff, L. & Wouters, P., eds (1997), *Proceedings of the Erasmus Workshop on Quantitative Approaches to Science & Technology Studies. May 21–24, 1996, Amsterdam (The Netherlands)*, Vol. 38.

Lorenz, C. (1987, 1994), *De constructie van het verleden. Een inleiding in de theorie van de geschiedenis*, Boom.

Luhmann, N. (1985), *Soziale Systeme*, Suhrkamp, Frankfurt ad Main.

Luhmann, N. (1992), *Die Wissenschaft der Gesellschaft*, Suhrkamp Taschenbuch.

Luukkonen, T. (1990), Citations in the rhetorical, reward and communications systems of science, PhD thesis, Acta Universitatis Tamperensis, ser A, vol. 285, University of Tampere, Tampere.

Luukkonen, T. (1997), Why has Latour's Theory of Citations been Ignored by the Bibliometric Community? Discussion of Sociological Interpretations of Citation Analysis, *in* Leydesdorff & Wouters (1997), pp. 27–39.

Lynch, M. & Woolgar, S. (1990), *Representation in scientific practice*, MIT Press, Cambridge.

MacRoberts, M. H. & MacRoberts, B. R. (1984), 'The negational reference: or the art of dissembling', *Social Studies of Science* **14**, 91–94.

MacRoberts, M. H. & MacRoberts, B. R. (1989), 'Problems of citation analysis: A critical review', *Journal of the American Society for Information Science* **40**(5), 342–349.

Marshakova, I. V. (1973), 'Bibliographic coupling system based on references (science citation index', *Nauch-Tekhn. Inform. Ser. 2 SSR* **2**, 3–8.

Marton, J. (1985), 'Obsolence or immediacy? evidence supporting price's hypothesis', *Scientometrics* **7**, 145–153.

Martyn, J. (1965), 'An examination of citation indexes', *Aslib Proceedings* **17**(6), 184–196.

Martyn, J. (1966), 'Citation indexing', *The Indexer* **5**(1), 5–15.

Maturana, H. R. & Varela, F. J. (1988), *The Tree of Knowledge*, New Science Library, Boston.

Merton, R. K. (1938), 'Science, technology and society in seventeenth-century england', *Osiris: Studies on the History and Philosophy of Science, and on the History of Learning and Culture* **IV**(2), 360–632.

Merton, R. K. (1941), 'Znaniecki's the social role of the man of knowledge', *American Sociological Review* **6**, 111–115.

Merton, R. K. (1973), *The normative structure of science*, University of Chicago Press, chapter 2.

Back to Top of Bibliography

Merton, R. K. (1977), The sociology of science: An episodic memoir, *in* Merton & Gaston (1977), chapter 1, pp. 3–141.

Merton, R. K. (1978), *Science, Technology and Society in Seventeenth-Century England*, Humanities Press, New Jersey.

Merton, R. K. & Gaston, J. (1977), *The sociology of science in Europe*, Southern Illinois University Press, Carbondale.

Miller, E. (1961), 'The national library of medicine index mechanization project', *Bulletin of the Medical Library Association* **49**(1), 1–96.

Moed, H. (1989), 'Bibliometric measurement of research performance and price's theory of differences among the sciences', *Scientometrics* **15**, 473–483.

Moed, H. & van Leeuwen, T. (1995), 'Improving the accuracy of isi's journal impact factors', *Journal of the American Society for Information Science* **46**, 461–467.

Molenaar, L. (1994), *Wij kunnen het niet langer aan de politici overlaten'. De geschiedenis van het Verbond van Wetenschappelijke Onderzoekers 1946-1980*, Elmar, Delft.

Mombers, C. M., van Venetië, R. & van Heeringen, A. (1983), 'Het "wetenschaps- en technologie-indicatoren" project'. RW 1517, Raadsvergadering 4 februari 1983, Agendapunt 9.

Mulkay, M., Potter, J. & Yearly, S. (1983), Why an analysis of scientific discourse is needed, Sage, London, pp. 171–203.

Nalimov, V. V. (1966), 'Kolichestvennye metody issledovaniya protsessa razvitiya nauki (quantitative methods of studying the development process of science)', *Voprosy filosfii (Problems of philosophy)* **12**, 38–47. Translated by Lilita Dzirkals, The Rand Corporation, July 1970.

Nalimov, V. V. (1992), 'V labirintakh bytiya. an interview by v. golavanov', *Literaturnava Gazeta* **21**.

Nalimov, V. V. & Multchenko, Z. M. (1969), *Naukometriya. Izucheniye Razvitiya Nauki kak Informatsionnogo Protsessa (Scientometrics: Studying science as an information process)*, Moscow, Nauka. English translation by the Foreign Technology Division, US Air Force Systems Command, 13 October 1971, microfilm, Washington DC.

Narin, F. (1969), Technology in retrospect and critical events in science. Illinois Institute of Technology Research. Report prepared for NSF, Washington DC.

Narin, F. (1976), *Evaluative Bibliometrics*, Computer Horizons Inc., Cherry Hill, N.J.

Newell, A. & Simon, H. A. (1990), Computer science as empirical enquiry: Symbols and search, *in* M. Boden, ed., 'The Philosophy of Artifical Intelligence', Oxford Readings in Philosophy, Oxford University Press, Oxford, pp. 105–133.

NHI (1984), 'Commentaar op het rapport van de Raad voor Advies van het Wetenschapsbeleid: "Prioriteiten in het gezondheidsondezoek"'. Letter to the Ministry of WVC.

Oberski, J. E. J. (1987), *Cocitaties clusteranalyse en natuurkunde*, RAWB, The Hague.

Oberski, J. E. J. (1988), *Handbook of Quantitative Studies of Science and Technology*, Elsevier Science Publishers, Amsterdam, chapter Some statistical aspects of Co-citation cluster analysis and a judgment by physicists, pp. 431–462.

O'Connell, J. (1993), 'Metrology: The creation of universality by the circulation of particulars', *Social Studies of Science* **23**, 129–173.

Odlyzko, A. M. (1995), 'Tragic loss or good riddance: The impending demise of traditional scholarly journals', *International Journal of Human-Computer Studies* . TO be obtained by email: 'send tragic.loss.long.ps' to netlib@research.att.com.

OenW (1966), 'Punten voor installatierede voor de Raad voor het Wetenschapsbeleid'. DGW 149411/HOW Archief nr. 129.

OenW (1981), 'Adviesaanvrage m.b.t. het gezondheidsonderzoek in nederland'. DGWB 18.898.

OenW (1985), 'Regeringsstandpunt RAWB-advies "Prioriteiten in het gezondheidsonderzoek"'. Letter to the RAWB, DGWB 34.505; RW 1880b.

OenW (1986*a*), 'Informatieplan hoofddirectie wetenschapsbeleid: Primaire informatievoorziening'. Internal memo.

OenW (1986*b*), 'Informatieplan hoofddirectie wetenschapsbeleid, versie 1'. Draft.

OenW (1986*c*), 'Memo betreft: Wbu '88'. Internal memo from Siskens, Broesterhuizen, Blaauw.

OenW (1986*d*), 'Voorstellen directie-pb nav nota nr. pb/1370'. Internal communication.

OenW (1987*a*), 'Agendapunten voor bespreking met RAWB/cbs, 16-6-1987'. Internal memo.

OenW (1987*b*), 'Besluitenlijst vergadering indicatorenwerkgroep dd. 4 augustus 1987'. Internal memo.

OenW (1987*c*), 'Besluitenlijstje werkgroep informatieplan 20-2-'87'. Internal memo.

OenW (1987*d*), 'Informatieplan hoofddirectie wetenschapsbeleid'.

OenW (1987*e*), 'Verslag van het gesprek over wb-trendanalyses op 16 juni 1987'. Internal memo.

OenW (1987*f*), 'Werkplan indicatoren'. Internal memo.

Ossowska, M. & Ossowski, S. (1936), 'The science of science', *Organon* **1**(1), 1–12. republished in: *Minerva*, vol. III, number 1 (Autumn 1964), pp. 72-82.

Porter, T. M. (1995), *Trust in numbers. The pursuit of objectivity in science and public life*, Princeton University Press.

Price, D. d. (1951), 'Quantitative measures of the development of science', *Archives Internationales d'Histoire des Sciences* **14**, 85–93. Presented at the VIe Congrès International d'Histoire des Sciences, Amsterdam, 1950.

Price, D. d. (1961), *Science since Babylon*, Yale University Press, New Haven.

Price, D. d. (1963), *Little Science, Big Science*, Columbia University Press, New York.

Price, D. d. (1965*a*), 'Networks of scientific papers', *Scientometrics* **149**, 510–515.

Price, D. d. (1965*b*), 'The scientific foundations of science policy', *Nature* **206**, 233–238.

Price, D. d. (1970), Citation measures of hard science, soft science, technology and non-science, D. C. Heath & Co., Lexington, Mass., pp. 3–22.

Price, D. d. (1978), 'Editorial statements', *Scientometrics* **1**, 7.

Price, D. d. (1979), The citation cycle, *in* 'The American Society for Information Science, 8th Mid-Year Meeting, May 16-19, 1979 Collected Papers'.

Price, D. K. (1962), 'The scientific establishment', *Proceedings American Philosophical Society* **106**(3).

PSAC (1958), 'Improving the availability of scientific and technical information in the united states. a report of the president's science advisory committee', The White House.

PSAC (1963), 'Science, government and information. a report of the president's science advisory committee', The White House January.

Raisig, L. M. (1960), 'Mathematical evaluation of the scientific serial', *Science* **131**, 1417–1419.

RAWB (1977), 'Bijlagen bij de raadsvergadering 6 mei 1977, agendapunt 5'. RW 968.

RAWB (1978*a*), 'De bevordering van wetenschapsonderzoek'. RW 1049.

RAWB (1978*b*), 'Voorlopige notitie wetenschapsonderzoek'. Bijlage bij RW 1049.

RAWB (1980*a*), 'Advies van de commissie wetenschapsdynamica aan de RAWB'. RW 1294.

RAWB (1980*b*), 'RAWB, verslag van de 160ste vergadering op 7 november 1980'. RW 1299.

RAWB (1982*a*), 'De RAWB in de vierde zittingsperiode. concept'. RW 1404, Raadsvergadering 5 febr. 1982, Agendapunt 5.

RAWB (1982*b*), 'RAWB, verslag van de 178ste vergadering van 4 juni 1982'. RW 1445.

RAWB (1983*a*), *Advies inzake de Prioriteiten in het gezondheidsonderzoek*, RAWB.

RAWB (1983*b*), 'Concept-advies 'prioriteiten in het gezondheidsonderzoek''. RW 1573, Raadsvergadering 1-7-1983, Agendapunt 5.

RAWB (1983*c*), 'Conceptadvies "prioriteiten in het gezondheidsonderzoek"'. RW 1563.

RAWB (1983*d*), 'Inleiding t.b.v. bijeenkomst met cvb universiteiten ter toelichting op gezondheidsadvies', RAWB archive, The Hague.

RAWB (1983*e*), 'RAWB, verslag van de 190ste raadsvergadering'. RW 1579.

RAWB (1983*f*), 'RAWB, verslag van de 194ste vergadering op 2 december 1983, te 10.00 uur'. RW 1635.

RAWB (1983*g*), 'Verslag van de 188e raadsvergadering, 3 juni 1983'. RW 1565.

RAWB (1984), 'Notitie t.a.v. de reacties op het RAWB-advies 'prioriteiten in het gezondheidsonderzoek''. RW 1722.

Rigter, H. (1980), 'Voorstel ten aanzien van project medisch-wetenschappelijk onderzoek (mwo)'. RW 1310, Raadsvergadering 9-1-1981, Agendapunt 5.

Rigter, H. (1981*a*), 'Informatie over het project 'prioriteiten medisch-wetenschappelijk onderzoek'. RW 1379, Raadsvergadering 6 nov. 1981, Agendapunt 8.

Rigter, H. (1981*b*), 'Omschrijving van het project medisch-wetenschappelijk onderzoek (MWO)'. RW 1319, Raadsvergadering 6 febr. 1981, Agendapunt 3.

Rigter, H. (1982), 'Beknopt voortgangsverslag van het project 'prioriteiten medisch-wetenschappelijk onderzoek''. RW 1442, Raadsvergadering 4 juni 1982, Agendapunt 6.

Rigter, H. (1983*a*), *De omvang en aard van het gezondheidsonderzoek in Nederland*, number 8 *in* 'Serie Achtergrondstudies', RAWB.

Rigter, H. (1983*b*), 'De prestaties van het nederlandse gezondheidsonderzoek in maat en getal'. RW 1536, Raadsvergadering 4 maart 1983, Agendapunt 5.

Rigter, H. (1983*c*), *De prestaties van het Nederlandse gezondheidsonderzoek. toepassing van een aantal wetenschapsindicatoren*, number 9 *in* 'Serie Achtergrondstudies', RAWB.

Rigter, H. (1983*d*), 'Het stellen van prioriteiten in het volksgezondheidsbeleid en het gezondheidsonderzoek. Achtergrondgegevens. Ruw concept.'. RW 1559.

Rigter, H. (1984), 'Haantje de voorste of lest best?', *Nederlands Tijdschrift voor Geneeskunde* **128**(27), 1286–1288.

Rigter, H. (1985). Letter to Wim Hutter; U2603/HR/slb.

Rigter, H. (1986), 'Evaluation of performance of health research in the netherlands', *Research Policy* **15**, 33–48.

Rip, A. (1996), 'The postmodern research system', Paper presented at the 1996 Progress Conference of the Department of Science and Technology Dynamics, University of Amsterdam, September 16-17 1996.

Rip, A. & Courtial, J.-P. (1984), 'Co-word maps of biotechnology: an example of cognitive scientometrics', *Scientometrics* **6**, 381–400.

Romein, J. (1976*a*), *Historische lijnen en patronen. Een keuze uit de essays*, Querido, Amsterdam.

Romein, J. (1976*b*), Theoretische geschiedenis, in *Historische lijnen en patronen. Een keuze uit de essays* (Romein 1976*a*), pp. 245–271.

Back to start of Bibliography

Romein, J. (1976*c*), Zekerheid en onzekerheid in de geschiedwetenschap, in *Historische lijnen en patronen. Een keuze uit de essays* (Romein 1976*a*), pp. 90–119.

Rouse, J. (1987), *Knowledge and Power: Toward a Political Philosphy of Science*, Cornell University Press, Ithaca, NY.

Sassower, R. (1995), *Cultural Collisions. Postmodern Technoscience*, Routledge, New York and London.

Schneiders, P. (1982), De Bibliotheek- en Documentatiebeweging 1880–1914; Bibliografische Ondernemingen rond 1900, PhD thesis, University of Amsterdam, Amsterdam, The Netherlands.

Schoenbach, U. H. (1956), 'Citation indexes for science', *Science* **123**, 61–62.

Schubert, A. & Maczelka, H. (1993), 'Cognitive Changes in Scientometrics during the 1980s, as Reflected by the Reference Patterns of its Core Journal', *Social Studies of Science* **23**, 571–581.

Scott, P., Richards, E. & Martin, B. (1990), 'Captives of controversy: The myth of the neutral social researcher in contemporary scientific controversies', *Science, Technology & Human Values* **15**(4), 474–494.

Seidel, A. H. (1949), 'Citation system for Patent Office', *Journal of the Patent Office Society* **31**, 554.

Shank, R. (1965), 'New concepts in indexing', *Bulletin of the Medical Library Association* **53**(3), 388–398.

Shannon, C. E. & Weaver, W. (1949), *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.

Shapin, S. & Schaffer, S. (1985), *Leviathan and the Air Pump: Hobbes, Boyle and the experimental life*, Princeton University Press, Princeton, NJ.

Shaw, R. R. (1949), 'The rapid selector', *The Journal of Documentation* **5**(3), 164–171.

Shrum, W. & Mullins, N. (1988), *Handbook of Quantitative Studies of Science and Technology*, Elsevier Science Publishers, Amsterdam, chapter Network analysis in the study of science and technology, pp. 107–133.

Small, H. (1973), 'Co-citation in the scientific literature: A new measure of the relationship between two documents', *Journal of the American Society for Information Science* **24**, 265–269.

Small, H. (1977), 'A co-citation model of a scientific specialty: A longitudinal study of collagen research', *Social Studies of Science* **7**, 139–166.

Small, H. G. & Griffith, B. C. (1974), 'The structure of scientific literature. i: Identifying and graphing specialties', *Science Studies* **4**, 17–40.

Small, H. & Sweeney, E. (1985*a*), 'Clustering the Science Citation Index using co-citations ii. mapping science', *Scientometrics* **8**, 321–340.

Small, H. & Sweeney, E. (1985*b*), 'Clustering the science citation index using co-citations I. a comparison of methods', *Scientometrics* **7**, 391–409.

Smith, L. C. (1981), 'Citation analysis', *Library Trends* pp. 83–106.

SONO (1985), 'Sono nr. 1'.

Spiegel-Rösing, I. (1977), 'Science studies: Bibliometric and content analysis', *Social Studies of Science* **7**, 97–113.

Spiegel-Rösing, I. & de Solla Price, D. (1977), *Science, Technology and Society: a cross-disciplinary perspective*, Sage, London.

Starchild, B. et al., eds (1981), *ISI Atlas of Science. Biochemistry and Molecular Biology 1978/80 including Minireviews of 102 Research Front Specialties*, ISI, Philadelphia, Pennsylvania, USA.

Steinbach, H. B. (1964), 'The quest for certainty: Science Citation Index', *Science* **145**, 142–143.

Stokes, T. D. & Hartley, J. A. (1989), 'Co-authorship & influence in specialties', *Social Studies of Science* **19**, 101–125.

Sullivan, D., White, D. H. & Barboni, E. J. (1977), 'Co-citation analyses of science: An evaluation', *Social Studies of Science* **7**, 223–240.

Thomassen, P. & Stanley, J. C. (1955), 'Uncritical citation of critized data', *Science* **121**, 610–611.

Tollebeek, J. (1996), *De ekster en de kooi*, Uitgeverij Bert Bakker.

Tukey, J. W. (1962), 'Keeping research in contact with the literature: Citation indices and beyond', *Journal of Chemical Documentation* **2**, 34–37. Presented before the Division of Chemical Literature, ACS National Meeting, Chicago, Ill., September 6, 1961.

Tukey, J. W. (n.d.*a*), The Citation Index and the information problem: Opportunities and research in progress. Statistical Techniques Research Group, Princeton University. Annual Report for 1962 under NSF Grant NSF-G-22108.

Tukey, J. W. (n.d.*b*), A Citation Index for Statistics: An introduction and a progress report. Xerox available in Garfield's Personal Archive.

van Bekkum, D. W. (1983), 'Kan RAWB oordeel over wetenschappelijk onderzoek vellen?', *NRC/Handelsblad* .

van Bueren, D. H. G. (1983*a*), 'Prioriteiten in het gezondheidsonderzoek. de raad van advies voor het wetenschapsbeleid en de moed om keuzen te maken', *Medisch Contact* **38**(51), 1597–1598.

van Bueren, H. G. (1983*b*), 'Om goed wetenschappelijk onderzoek te handhaven is oordeel onontbeerlijk', *NRC/Handelsblad* .

van Bueren, H. G. (1984), 'Prioriteiten in het gezondheidsonderzoek'. RW 1730, Letter to Mrs. Bensing, NIH.

van der Meer NZI, D. D. (1984). Letter to the Ministry of WVC.

Van der Meulen, B. J. R. (1992), Evaluation Processes in Science: the Construction of Quality by Science, Government and Industry, PhD thesis, University of Twente.

van Duyne, W. M. J. (1984), 'Commentaar op het advies van de RAWB: "Prioriteiten in het gezondheidsondezoek"'. Letter to the Ministry of WVC.

van Es, J. C. (1983), 'Wetenschapsbeleid', *Medisch Contact* **38**(43), 1351.

van Heeringen, A. (1979), Het effect van arbeidsmobiliteit en leeftijd op de produktiviteit van onderzoekers. concept projectvoorstel. Bijlage bij GWO Nieuwsbrief nr. 56.

van Heeringen, A. (1982), 'Verslag van een bezoek aan de Verenigde Staten van 14 juni - 30 juni 1982'. RW 1518.

van Heeringen, A. & Langendorff, A. N. M. (1988), *Wetenschaps- en technologie-indicatoren 1988*, RAWB, The Hague.

van Heeringen, A., Mombers, C. & van Venetië, R. (1984), *Wetenschaps- en Technologie Indicatoren 1983*, number 11 *in* 'Serie Achtergrondstudies RAWB', RAWB.

van Onderwijs en Wetenschappen, M. (1981), 'Adviesaanvrage m.b.t. het gezondheidsonderzoek in Nederland'. DGWB 18.898 / RW 1364, Raadsvergadering 2 okt. 1981, Agendapunt 3.

van Raan, A. F. J. (1987*a*), 'Onderzoekprogramma 'Trends in het Nederlandse Speur- en Ontwikkelingswerk"'.

van Raan, A. F. J., ed. (1988), *Handbook of Quantitative Studies of Science and Technology*, Elsevier Science Publishers, Amsterdam.

van Raan, A. F. J., Soete, L., Beelen, E., de Bruin, R. E., Moed, H. F., Nederhof, A. J., Noyons, E. & Negenborn, P. (1994), *Het Nederlands Observatorium van Wetenschap en Technologie. Wetenschaps- en Technologie-Indicatoren 1994*, Wetenschaps- en Technologie-Indicatoren, Ministerie van Onderwijs en Wetenschappen, Zoetermeer.

van Raan, T. (1987*b*), 'Voorstel tot een vooronderzoek in het kader van een onderzoekprogramma 'trends in het nederlands speur- en ontwikkelingswerk''.

van Raan, T. (1987*c*), 'Wensen RAWB mbt WTI-programma april-september 1987'. Letter to Arie van Heeringen, RAWB.

van Rooyen, N., Doorsma, D. M. & Eikelenboom, P. (1983), 'Citaties onmisbaar voor goed onderzoek', *NRC/Handelsblad* pp. Bijvoegsel Wetenschap en Onderwijs pp. 1–2.

Vladutz, G. E., Nalimov, V. V. & Stiazkhin, N. I. (1959), 'Nauchnaya informatsiya kak zadacha kibernetika (scientific and technical information as a task of cybernetics)', *Uspekhi fizicheskikhnauk (Achievements in the physical sciences)* **69**(1).

Vucinich, A. (1982), 'Soviet marxism and the history of science', *The Russian Review* **41**, 123–143.

W, O. . (1991), Beleidsnotitie indicatoren. wetenschapsbudget 1991 bijlage c, Technical report, Ministerie van Onderwijs en Wetenschappen.

Welt, I. D. (1956), 'Subject indexing in a restricted field', *Science* **123**, 723–724.

Werskey, G. (1978), *The Visible College. A collective biography of British scientists and socialists of the 1930s*, Allen Lane, London.

Westheimer, F. (1965), Chemistry: Opportunities and needs, Technical report, National Academy of Sciences — National Research Council.

Whitley, R. (1984), *The intellectual and social organization of the sciences*, Oxford University Press.

Wil (1991), *Collins Cobuild English Language Dictionary*.

Woolgar, S. (1988*a*), *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*, Sage, London.

Woolgar, S. (1988*b*), *Science: The very idea.*, Tavistock, London.

Woolgar, S. (1991), 'Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy', *Science and Public Policy* **18**, 319–326.

Wouters, P. (1992*a*), 'Cijfers voor beleid, vragen voor de wetenschap', *Kennis & Methode* (1), 128–47.

Wouters, P. (1992*b*), 'Een nieuw teken aan de wand. De uitvinding van de Science Citation Index 1948-1964', *Kennis & Methode* **XVI**(4), 313–336.

Wouters, P. (1996*a*), 'Een wandelingetje in een witte bloedcel', *Elsevier* pp. 98–102.

Wouters, P. (1996*b*), 'Weg met zelfbeheer wetenschap. Minister Jo Ritzen vindt dat de politiek meer zeggenschap moet krijgen over het wetenschappelijk onderzoek', *Elsevier* **52**, 99.

Wouters, P. (1997*a*), Citation cycles and peer review cycles, *in* Leydesdorff & Wouters (1997), pp. 39–55.

Wouters, P. (1997*b*), 'Tegenoffensief. Elsevier Science, de grootste uitgever van wetenschappelijke tijdschriften ter wereld, start eindelijk een eigen elektronisch netwerk.', *Elsevier* **53**(11), 92–93.

Wouters, P. (1998*a*), 'Boeken tellen niet mee', *Hypothese* pp. 7–9.

Wouters, P. (1998*b*), 'The signs of science', *Scientometrics* **41**(1–2), 225–241.

Wouters, P. (1999*a*), 'Beyond the Holy Grail: From citation theory to indicator theories', *Scientometrics* **forthcoming**.

Wouters, P. (1999*b*), 'The creation of the SCI', *ASIS Monography Series* .

Wouters, P. F. (1996*c*), 'Cyberscience', *Kennis en Methode* **XX**(2), 155–186.

Wouters, P. & Leydesdorff, L. (1994), 'Has Price's dream come true: Is scientometrics a hard science?', *Scientometrics* **31**(2), 193–222.

Ziman, J. (1979), *From Parameters to Portents—and Back*, in Elkana et al. (1979), chapter 11, pp. 261–283.

Zirkle, C. (1954), 'Citation of fraudulent data', *Science* **120**, 189–190.

Zuckerman, H. A. (1971), 'Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity', *The American Journal of Sociology* pp. 276–291. Publication no. A-489 of the Bureau of Applied Social Research, Columbia University.