Big Data pilot project on the use of Wikipedia page views on World Heritage Sites

This section contains first experimental results of a pilot project intended to explore the potential of Big Data for official statistics. The results are based on an analysis of Wikipedia page views of World Heritage Sites.

Whenever people use computer-mediated systems, they unintentionally leave digital traces of their activity. These traces are increasingly becoming a relevant source of information for official statistics, in particular due to the potential they offer for improving the timeliness, the detail and, in some cases, the accuracy of the statistics.

For the purpose of this project Wikipedia page view counts constituted the main data source.

Since its launch in 2001, Wikipedia has grown to hold a total of 39 million articles in 246 languages (1). It is widely used, recording 21 million page views per hour (in April 2016) (2).

According to the Community survey on ICT usage in households and by individuals, in 2015 45 % of individuals between the ages of 16 and 74 living in the EU consulted wikis to obtain knowledge (e.g. Wikipedia). The proportion was 66 % amongst 16 to 24 year olds. Such widespread usage means that data about people's use of Wikipedia could potentially be a relevant big data source for producing official statistics.

The pilot project used data on the number of page views per month for all articles in 31 language versions of Wikipedia (3). The data used is made publicly available by the Wikimedia Foundation. The version of Wikipedia designed for use on mobile devices was not included in the number of views.

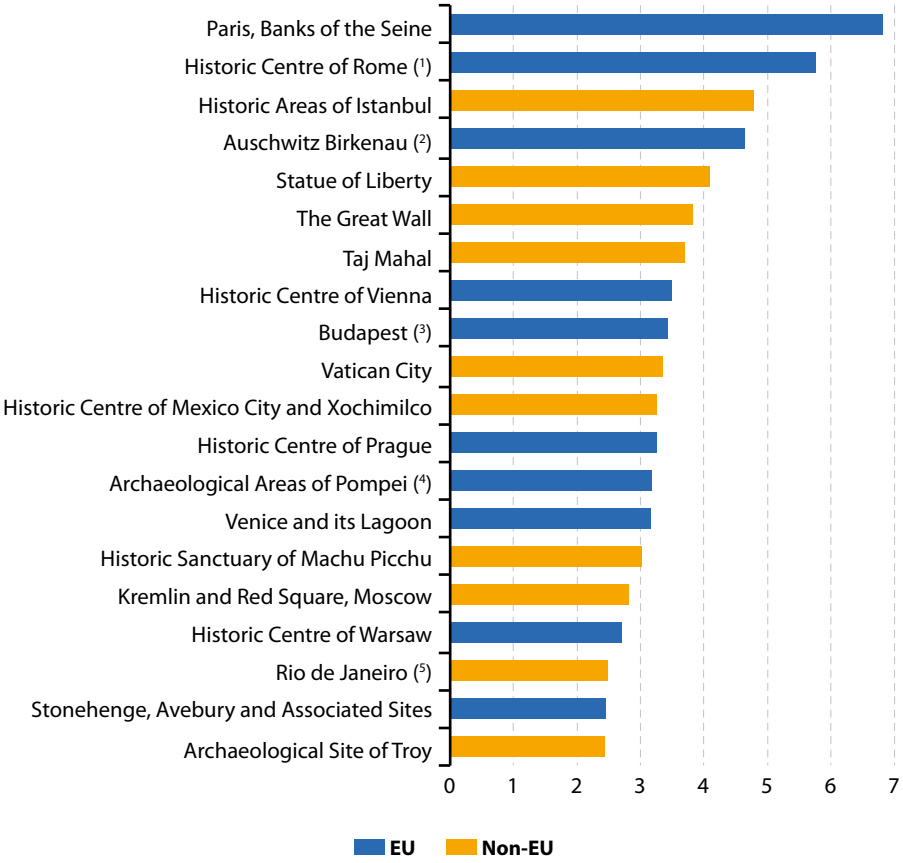
Wikipedia articles were selected for each of the 1 031 World Heritage Sites included on UNESCO's list in 2015. The initial selection was based on the categorisation feature in the English version of Wikipedia. The articles were linked to the corresponding heritage site on the basis of the information from the info box entitled 'World heritage site'. Other language versions of the same articles where then selected, where available. Data analysis was based on the number of page views of around 50 000 articles.

This pilot was run in the context of the Big Data Sandbox, an international collaboration project sponsored by the High-Level Group for the Modernisation of Official Statistics, set up by the Conference of European Statisticians. It involved, besides Eurostat, several national statistical institutes and other international statistical bodies.

More information can be found on the Eurostat website.

(1) From https://en.wikipedia.org/wiki/Special:Statistics consulted on 20 May 2016.
(2) From https://stats.wikimedia.org/EN/TablesPageViewsMonthlyCombined.htm, consulted on 20 May 2016.
(3) 31 Language versions cover 24 official EU languages and as well Icelandic, Macedonian, Norwegian, Russian, Albanian, Serbian and Turkish.

Figure 4: Top 20 World Heritage Sites in number of page views of related Wikipedia articles, 2015
(number of views in million)



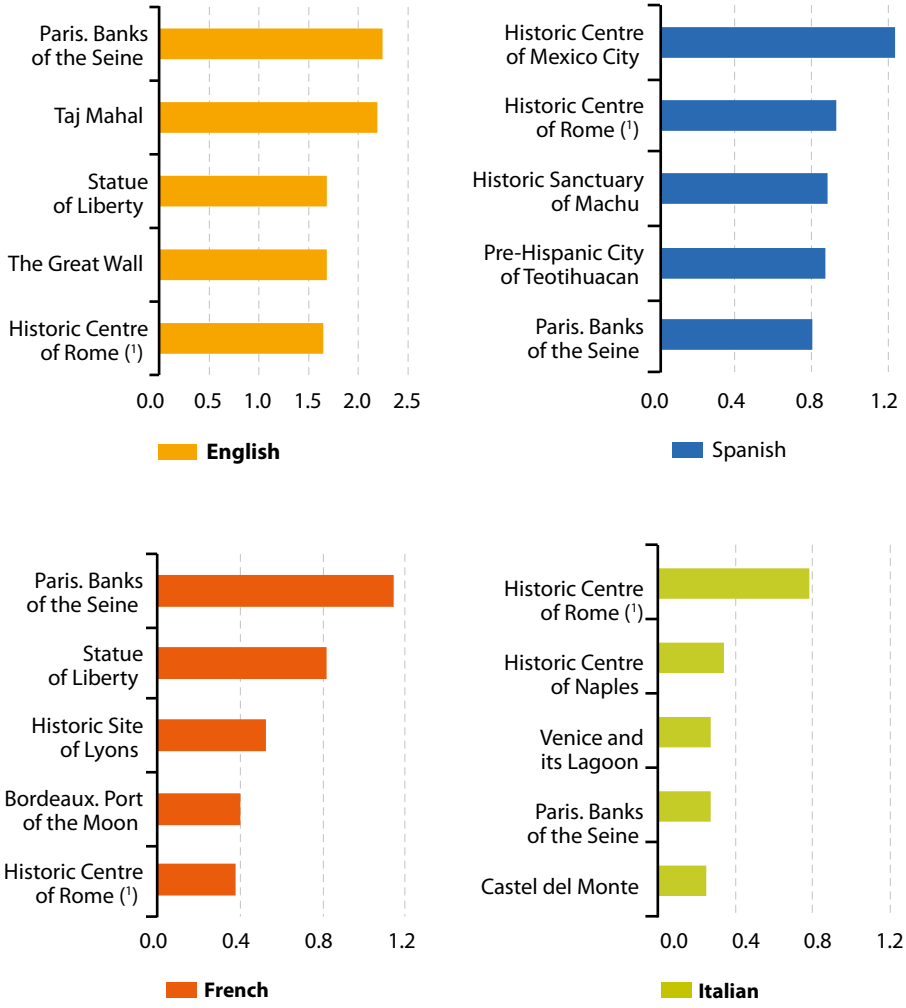
Note: experimental statistics based on the pilot project.

For proper graphic visualisation, some names have been shortened. See below the full names:

(1) Historic Centre of Rome, the Properties of the Holy See in that City Enjoying Extraterritorial Rights and San Paolo Fuori le Mura. Includes properties of the Holy See.
(2) Auschwitz Birkenau, German Nazi Concentration and Extermination Camp (1940-1945).
(3) Budapest, including the Banks of the Danube, the Buda Castle Quarter and Andrássy Avenue.
(4) Archaeological Areas of Pompei, Herculaneum and Torre Annunziata.
(5) Rio de Janeiro: Carioca Landscapes between the Mountain and the Sea.

Source: Eurostat and UNESCO, World Heritage List

Figure 5: Top 5 World Heritage Sites in number of page views of related Wikipedia articles by language, 2015
(number of views in million)



Note: experimental statistics based on the pilot.

(1) For proper graphic visualisation, this name has been shortened. The full name is Historic Centre of Rome, the Properties of the Holy See in that City Enjoying Extraterritorial Rights and San Paolo Fuori le Mura. Includes properties of the Holy See.

Source: Eurostat and UNESCO, World Heritage List