ICBO: International Conference on Biomedical Ontologies · Buffalo, NY, USA
Working with Multiple Biomedical Ontologies Workshop, July 26, 2011

Towards a Body Fluids Ontology: A Unified Application Ontology for Basic and Translational Science

Jiye Ai[1], Mauricio Barcellos Almeida[2], André Queiroz de Andrade[3], Alan Ruttenberg[4], David Tai Wai Wong[1], Barry Smith[5]

[1]University of California at Los Angeles, School of Dentistry and Dental Research Institute, CA, USA
[2]Federal University of Minas Gerais, School of Information Science, Belo Horizonte, MG, Brazil
[3]Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[4]State University of New York at Buffalo, Clinical and Translational Data Exchange, Buffalo, NY, USA
[5]State University of New York at Buffalo, Philosophy Department, Buffalo, NY, USA

Abstract. We describe the rationale for an application ontology covering the domain of human body fluids that is designed to facilitate representation, reuse, sharing and integration of diagnostic, physiological, and biochemical data. We briefly review the Blood Ontology (BLO), Saliva Ontology (SALO) and Kidney and Urinary Pathway Ontology (KUPO) initiatives. We discuss the methods employed in each, and address the project of using them as starting point for a unified body fluids ontology resource. We conclude with a description of how the body fluids ontology initiative may provide support to basic and translational science.

Keywords: body fluids, ontology, saliva, human blood.

1 Introduction

Body fluids are liquids that are excreted or secreted from and inside the bodies of organisms. We here focus on the case of human organisms, and provide a preliminary assay of the scope and purpose of an application ontology covering the domain of body fluids in both healthy and diseased human organism.

Only a small fraction of the human body fluids have been included thus far in the Foundational Model of Anatomy (FMA) [1]. These are listed, with their definitions, in Table 1.

Body Fluid	Definition (FMA)
blood	Portion of body substance that consists of plasma and blood cells.
breast milk	Portion of secreted substance produced by the mammary gland.
chyme	Ingested food admixed with gastric secretions contained in the stomach.
endolymph	Transudate contained within the membranous labyrinth.
mucus	Portion of secreted substance produced by a mucous gland or goblet cell.
perilymph	Transudate contained in the osseous labyrinth outside the membranous labyrinth.
plasma	Body substance in liquid state contained in the lumen of arterial and venous trees, blood capillary and the cardiac chambers; constitutes the liquid phase of blood.
semen	Portion of body fluid suspension that consists of spermatozoa and seminal plasma.
sweat	Secretion produced by a sweat gland.
seminal fluid	Portion of secreted substance produced prostatic glands, bulbourethral glands or the seminal vesicles.
serum	Body substance derived from plasma by the elimination of fibrinogen.
tear	Portion of secreted substance produced by the lacrimal gland.
urine	Excretion in liquid state processed by the kidney.

Table 1. List of body fluid types currently represented in FMA

Further human body fluid types which we have identified in our researches thus far, and which will be submitted to the FMA for inclusion in due course, include:

- bile
- aqueous humour
- cerebrospinal fluid
- cerumen
- colloidal body substance
- deferent duct fluid
- epididymal duct fluid
- esophageal secretion
- follicular fluid
- gastric juice
- gingival fluids
- intestinal secretion
- intraocular fluid
- lymph
- epithelial lining fluid
- lung lining fluid
- menstrual fluid
- pancreatic juice
- renal filtrate
- rete testis fluid
- saliva
- sebum
- seminiferous tubule fluid
- serous fluid
- synovial fluid
- tissue fluid
- vaginal lubrication
- vitreous humour

Body fluids in a broader sense include also fluids such as liquid feces which exist in the organism in a state where they are dissolved in water. They include also fluids that result from procedures such as bronchial lavage. However, because the FMA deals only with body fluids present in the 'canonical' human body, it does not include terms representing fluids which arise in cases of disease and in the performance of clinical procedures.

There is a continuous flow of body fluids throughout the body. They function as vehicles to carry oxygen, nutrients, waste, hormones and other signal molecules and immune sensors and effectors between the body's different compartments. Body fluids serve also as transporters for pharmaceutical substances. Approximately 60 percent of the human body consists of fluids. Portions of fluid within the interior of the cell are called intracellular fluids. All other fluids are extracellular, and it is these that we focus on here, and primarily on those extracellular fluids that are of value for diagnostic purposes.

Body fluids are present in the body in various combinations and in various proportions. An excess or shortage of a given body fluid in a given compartment or conduit can be a symptom or a cause of disease. Many diseases can affect body fluids in their turn, and the latter can thus serve as a diagnostic indicator of the former. This holds for some cancers [2, 3], kidney diseases [4], inflammatory diseases [5], and metabolic diseases [6]. Certain body fluids, above all blood, urine and saliva, provide advantages with regard to disease diagnosis and prognosis, primarily due to low invasiveness, minimal cost, and easy sample collection and processing [7, 8].

With the recent advances in biomedical technology, there has been an escalating need for formal tools that can facilitate effective and efficient representation, reuse, sharing and integration of diagnostic data. Scientists are increasingly recognizing the value of ontologies in this connection. An ontology provides a controlled vocabulary that can be shared by investigators in different fields, who can draw on the ontology's logical definitions to ensure that terms are used with common meanings. In addition the ontology can support quality control in data entry and allow algorithmic reasoning on data annotated using its terms.

The ontology-based approach will function successfully, however, only if the ontologies themselves are developed in tandem with each other in such a way as to ensure cross-domain consistency and to eliminate the sorts of redundancy in vocabulary creation which have traditionally arisen where domains overlap. To this end, a distinction needs to be drawn between reference and application ontologies [9]. The former correspond in medicine to the basic biomedical sciences such as anatomy and physiology, the latter to clinical specialisms and sub-specialisms, for example to pediatric surgery or radiation oncology. Just as the clinical specialisms draw on the content and results of the basic sciences, so application ontologies will need to draw on more basic feeder ontologies such as the aforementioned FMA, and others such as the Chemical Entities

of Biological Interest (ChEBI) [10], Protein Ontology (PRO) [11], Gene Ontology (GO) [12] and the Cell-Type Ontology (CT) [13].

Our proposed application ontology BFLO is designed to meet the need for terminology support in the domain of research on bio-fluids. As an application ontology, it will be built primarily out of terms deriving from other more foundational ontologies and terminologies. We shall draw most importantly on the FMA as our overarching anatomy framework [1]. There a *Portion of body fluid* is defined as:

> A portion of body substance that consists of a mixture of fluid, solutes and particles.

FMA: *Portion of body substance* is defined in turn as:

> Material anatomical entity in a gaseous, liquid, semisolid or solid state, with or without the admixture of cells and biological macromolecules; produced by anatomical structures or derived from inhaled and ingested substances that have been modified by anatomical structures.

All terms representing types of body fluid in our ontology will be treated as children of FMA: *Portion of body fluid*.

We will draw on other resources, including the SNOMED CT vocabulary of clinical terms, which includes terms relevant to the domain of body fluids such as *Body fluid (substance)*, *Origin of fluid (attribute)*, and *Body fluid retention (disorder)*. Body fluid (substance) is asserted in the SNOMED CT concept hierarchy to be both a Body substance and Liquid substance.

Apart from the FMA, the most important ontologies employed by the BFLO in its current alpha version are the Blood Ontology (BLO), the SALO Ontology (SALO) the Kidney and Urinary Pathway Ontology (KUPO).

1.1 Blood Ontology (BLO)

The BLO (http://mbaserver.eci.ufmg.br/BLO-wiki/) [14] is a controlled vocabulary designed for use in annotating and organizing data about blood, including data pertaining to:

- blood transfusions (for example, donation process control)
- hematology (for example, immunologic basis)
- blood derivative products (for example, frozen plasma)
- the content of regulatory documentation (for example, regulations under the Food and Drug Administration)
- the associated regulatory processes (for example tests of blood quality).

BLO is being created to serve the exploration and aggregation of information relevant to scientific research and to human blood manipulation. BLO is constructed on the basis of well-founded ontological principles in such a way that it can support interoperability with OBO Foundry ontologies such as the GO and the PRO. BLO corresponds to a set of interrelated ontologies, each addressing a group of relevant issues in the field of hematology and blood transfusion.

As concerns terms for diagnostic processes, BLO will rely on publications reporting the results of research on blood-transmitted diseases, for example, HIV-1/2, hepatitis B and C, Chagas Disease, and syphilis. In addition, the ontology will draw on terms used in research on hemophilias, Von-Willebrand disease and Sickle Cell Anemias, and on HTLV.

1.2 Saliva Ontology (SALO)

The SALO [15] is a consensus-based controlled vocabulary of terms and relations dedicated to the salivaomics domain and to saliva-related diagnostics. Like BLO, SALO follows the principles of the OBO Foundry.

The protein terms in the SALO are derived from the corresponding sections of the PRO (http://purl.org/obo/owl/PRO). The SALO is a component of an ontology-based framework for a Salivaomics Knowledge Base (SKB; http://www.skb.ucla.edu/ website) that is a data repository, management system and web resource constructed to support human salivary proteomics, transcriptomics, miRNA, metabolomics and microbiome research currently being assembled in the Dental Research Institute at the UCLA School of Dentistry [15].

1.3 Kidney and Urinary Pathway Ontology (KUPO)

The KUPO (http://www.e-lico.eu/public/kupo/kupo.owl) [16] is an ontology that describes kidney and urinary anatomy, the cells in the associated organs and tissue, and the gene products in those cells and their functional attributes and cellular components and associated pathologies. It also contains terms relating to transcriptomics and proteomics experiments in the KUPO domain. Fluid-related terms in KUPO include: urea transport, small muscle layer of renal vein, rennin secretion into blood stream, and disorder of ureter.

2 Method

All body fluids share certain features in common, including relations to biological process represented in the GO, to proteins represented in the PRO, and to cell types represented in the CT. Given the large variety of different types of body fluid, however, these common features must be specialized to each individual case. In developing BFLO we accordingly employ a novel method for ontology development, which we call the method of generalization and specialization.

In the simplest case, BLO and SALO are taken as an input, and corresponding terms in the two ontologies, such as *portion of blood* and *portion of saliva* are aligned in light of their common Basic Formal Ontology (BFO) [17] category (http://www.ifomis.org/bfo/). The aligned terms are then joined to form a new ontology term as output, for example: *portion of blood*, *portion of saliva* → *portion of body fluid*. The MIREOT [9] guidelines are followed at every stage when terms are imported into BFLO from other ontologies.

3 Results and Discussion

Figure 1 displays a fragment of the BFLO illustrating fundamental entities such as water, protein, DNA, RNA, cell, organ, etc.

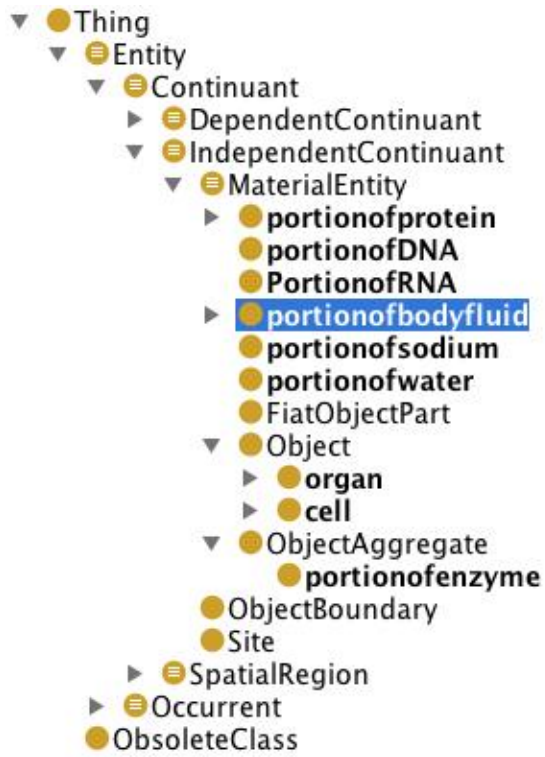

Figure 1. Fragment of BFLO

Application ontologies may be created either because reference ontologies are too large to be of effective service in specific, narrowly focused projects, or because they are too narrow in scope. We might, for example, be interested in data associated with physiological models of body fluid exchange [18], or in data which results from comparing the utility of different body fluids (for example saliva, blood and urine) for diagnostic purposes in relation to multiple different sorts of diseases. An application ontology such as BFLO can then help us to focus on the relevant content of the reference ontologies and to conjoin the corresponding fragments together in a way that helps us to address cross-domain issues.

Another set of examples of potential uses of BFLO concern comparisons of the diagnostic value of different body substances where the results of tests employing one substance point to the need for tests using some other substance.

To build an adequate representation of a biological phenomenon we need precise information about the biological components involved on several different levels of granularity. Proteomics researchers [19] have mapped the similarities and differences in

protein composition in plasma and saliva samples. They have shown, for instance, that imunoglobulins present in saliva and plasma overlap in a way that suggests leakage from plasma into saliva. Given the convenience of saliva sampling such leakage could provide the possibility of more expedient testing for antibodies. At the same time, it may be that we can use information about known biomarkers in blood to make inferences to the presence of as yet unknown biomarkers in saliva [20]. Integration of blood and saliva data could thus be advanced in useful ways through the creation of the unified BFLO framework. The representation would need to take into account also the fact that some proteins have different behaviors in special conditions (such as diseases), and thus proper functional annotation is essential if an ontology-based representation of data is to support prediction.

4 Conclusion and Future Directions

The full understanding of the physiology and of the body requires the use of data relating not merely to body fluids taken singly, but also in combination with other body fluids, and of course with other anatomical entities.

The techniques and applications sketched in the foregoing are just the first steps towards a truly useful BFLO. Close coordination with the OBO Foundry, consolidation of a common framework, and exploration of potential collaborations with KUPO, as well as expansion to other representative fluids, are the next steps in which we intend to move forward.

The growth of research on body fluids is raising important challenges for the field of biomedical ontologies, bringing the demand for a resource that can accelerate the consistent representation, organization and manipulation of body fluid data. We accordingly believe that a resource like the proposed BFLO has the opportunity to advance both basic and translational science.

Acknowledgments

This work was supported by US National Institutes of Health grants U01DE016275, U01DE017790 and U54HG004028.

References

1. Rosse, C., Mejino, J.L., Jr.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform 36, 478-500 (2003)
2. Gao, K., Zhou, H., Zhang, L., Lee, J.W., Zhou, Q., Hu, S., Wolinsky, L.E., Farrell, J., Eibl, G., Wong, D.T.: Systemic disease-induced salivary biomarker profiles in mouse models of melanoma and non-small cell lung cancer. PLoS One 4, e5875 (2009)
3. Wulfkuhle, J.D., Liotta, L.A., Petricoin, E.F.: Proteomic applications for the early detection of cancer. Nat Rev Cancer 3, 267-275 (2003)
4. Israni AK, e.a.: Laboratory assessment of kidney disease: Clearance, urinalysis, and kidney biopsy. Saunders Elsevier, (2008)
5. Young, B., Gleeson, M., Cripps, A.W.: C-reactive protein: a critical review. Pathology 23, 118-124 (1991)
6. Burton, B.K.: Inborn errors of metabolism in infancy: a guide to diagnosis. Pediatrics 102, E69 (1998)
7. Greer, J.P.e.a.: Wintrobe's clinical hematology. Lippincott Williams & Wilkins, (2008)
8. Veenstra, T.D., Conrads, T.P., Hood, B.L., Avellino, A.M., Ellenbogen, R.G., Morrison, R.S.: Biomarkers: mining the biofluid proteome. Mol Cell Proteomics 4, 409-418 (2005)
9. Courtot, M., Gibson, F., Lister, A.L., Malone, J., Schober, D., Brinkman, R.R., Ruttenberg, A.: MIREOT: The minimum information to reference an external ontology term. Appl. Ontol. 6, 23-33
10. de Matos, P., Alcantara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S., Steinbeck, C.: Chemical Entities of Biological Interest: an update. Nucleic Acids Res 38, D249-254 (2010)
11. Natale, D.A., Arighi, C.N., Barker, W.C., Blake, J., Chang, T.C., Hu, Z., Liu, H., Smith, B., Wu, C.H.: Framework for a protein ontology. BMC Bioinformatics 8 Suppl 9, S1 (2007)
12. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-29 (2000)
13. Bard, J., Rhee, S.Y., Ashburner, M.: An ontology for cell types. Genome Biol 6, R21 (2005)

14. Almeida MB., S.B., Proietti ABC. Coelho KC.: The Blood Ontology: organizing the information in the domain of the human blood.
15. Ai, J., Smith, B., Wong, D.T.: Saliva Ontology: an ontology-based framework for a Salivaomics Knowledge Base. BMC Bioinformatics 11, 302 (2010)
16. Simon Jupp, J.K., Joost Schanstra and Robert Steven.: Developing a Kidney and Urinary Pathway Knowledge Base. Bio-ontologies SIG 2010, Boston, USA (2010)
17. Grenon, P., Smith, B., Goldberg, L.: Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Technol Inform 102, 20-38 (2004)
18. Thomas, S.R., Baconnier, P., Fontecave, J., Francoise, J.P., Guillaud, F., Hannaert, P., Hernandez, A., Le Rolle, V., Maziere, P., Tahi, F., White, R.J.: SAPHIR: a physiome core model of body fluid homeostasis and blood pressure regulation. Philos Transact A Math Phys Eng Sci 366, 3175-3197 (2008)
19. Yan, W., Apweiler, R., Balgley, B.M., Boontheung, P., Bundy, J.L., Cargile, B.J., Cole, S., Fang, X., Gonzalez-Begne, M., Griffin, T.J., Hagen, F., Hu, S., Wolinsky, L.E., Lee, C.S., Malamud, D., Melvin, J.E., Menon, R., Mueller, M., Qiao, R., Rhodus, N.L., Sevinsky, J.R., States, D., Stephenson, J.L., Than, S., Yates, J.R., Yu, W., Xie, H., Xie, Y., Omenn, G.S., Loo, J.A., Wong, D.T.: Systematic comparison of the human saliva and plasma proteomes. Proteomics Clin Appl 3, 116-134 (2009)
20. Spielmann, N., Wong, D.T.: Saliva: diagnostics and therapeutic perspectives. Oral Dis 17, 345-354 (2011)