

Special Focus Document from the ELSA Consultation meeting, September 29, 2009

Biomedical Terminologies and Ontologies

Enabling Biomedical Semantic Interoperability and Standards in Europe

Bernard de Bono*(EMBL-EBI), **Mathias Brochhausen**(IFOMIS), **Sybo Dijkstra**(Philips), **Dipak Kalra**(CHIME), **Stephan Keifer**(IBMT), **Barry Smith**(Uni of Buffalo)

*Correspondence to: EMBL-EBI, Wellcome Trust Genome Campus, Cambridge CB10 1SD (bdb@ebi.ac.uk)

“Real international collaboration only began when electrical telegraphy became an important instrument of communication. Again, as in the case of the visual telegraph, at first there were only national networks...whenever a line reached a national frontier, there it ceased. This was as much due to the different systems employed, each having its own code vocabulary...”
ITU-T Report: 50 Years of Excellence (20/07/2006) [<http://www.itu.int/ITU-T/>]

In the management of biomedical data, vocabularies such as ontologies and terminologies (O/Ts) are used for (i) domain **knowledge representation** and (ii) **interoperability**. The knowledge representation role supports the automated reasoning on, and analysis of, data annotated with O/Ts. At an interoperability level, the use of a communal vocabulary standard for a particular domain is essential for large data repositories and information management systems to communicate consistently with one other. Consequently, the interoperability benefit of selecting a particular O/T as a standard for data exchange purposes is often seen by the end-user as a function of the number of applications using that vocabulary (and, by extension, the size of the user base). Furthermore, the adoption of an O/T as an interoperability standard requires confidence in its stability and guaranteed continuity as a resource.

Unregulated practices in vocabulary development and deployment impact the interoperability benefit of biomedical vocabularies, for example:

- i. In biological research, new O/Ts are typically developed to improve the knowledge representation capabilities for a particular domain of study. While competition among an increasing number of overlapping vocabularies stimulates further creative development in the field, the uncoordinated adoption of different O/Ts for the same domain also erodes into their collective interoperability benefit.
- ii. While scientific groups are becoming increasingly familiar with tools and techniques to create vocabularies, few are prepared for the level of resource commitment and community engagement required to sustain the (i) dissemination of their work to improve the interoperability benefit, (ii) maintenance of their vocabulary in a robust (and backwards-compatible) manner, (iii) adoption of the vocabulary by the community, (iii) provision of support to user feedback and feature requests, and (iv) co-ordination of development with complementary vocabularies in neighbouring domains.
- iii. Different vocabulary licensing conditions offered to academic, clinical and industry users may affect the universal adoption of O/Ts. For example, such restrictions may interfere with community-wide interoperability operations involving O/Ts that are not licensed for public use.

The presence of multiple overlapping O/Ts represents one aspect of the interoperability problem. At the other extreme is, of course, the complete lack of:

- iv. viable O/Ts to satisfy the demand in a particular field,
- v. awareness of the O/T benefits,
- vi. well-defined data structures that provide a proper context for O/T terms (term context is crucial in electronic health records as it allows entries to be linked chronologically as a means to trace the evolution of a patient's health status), as well as
- vii. a validated business model that motivates and rewards O/T adoption.

Achieving widespread semantic interoperability in key areas such as healthcare is a complex task¹. In view of the central role O/Ts play in the development of electronic health record (EHR) interoperability standards, the ELSA effort should invest in the active governance of biomedically-relevant vocabularies in order to assist with the challenges (i. to vii.) listed above. In particular, ELSA should seek to support research into a biomedical semantic interoperability standards development body that draws upon the following examples:

1. The Open Biomedical Ontology Foundry initiative that is driven by the knowledge representation and interoperability needs of a wide spectrum of the biological community in an effort to co-ordinate the coherent development and quality control of interoperable ontologies [<http://www.obofoundry.org/>];
2. ISO/IEEE/CEN 11073 Health Informatics Committee for the joint standards addressing the interoperability of medical devices [http://grouper.ieee.org/groups/detailed_index.html - EMB:11073];

¹ See: http://ec.europa.eu/information_society/activities/health/docs/publications/2009/2009semantic-health-report.pdf

3. ISO/EN 13606 Electronic Health Record Communication standard which defines models and interfaces for EHR interoperability and formal approaches for systematising clinical data structures [http://www.iso.org/iso/catalogue_detail.htm?csnumber=40784];
4. EMEA's Committee for Medicinal Products for Human Use (CHMP). This committee approves scientific advice given on the drug development process, as well as assessing whether a medicinal product has a positive risk benefit for a given indication (e.g. a disease state). Furthermore it prepares EMEA's opinion for the EC's marketing authorisation (initially for a fixed period of time) for a new product, as well as assessing significant modifications through the lifecycle of a product. It also issues expert guidelines on the investigational approach to drug development in particular disease states [<http://www.emea.europa.eu/htms/general/contacts/CHMP/CHMP.html>].

The ELSA initiative should therefore address directly challenges **i.** to **vii.** through a study of methodologies that are relevant to a body that supports the development of O/T-based semantic interoperability standards by:

- 1) Advising on **funding priorities** for new vocabularies (e.g. target indications (domains), knowledge representation requirements);
- 2) Issuing **guidelines on O/T development** (e.g. dealing with overlap, interoperability and alignment between O/Ts, avoidance of redundancy and ambiguity, lifecycle management, maintenance, versioning)
- 3) Issuing **guidelines on O/T deployment** (e.g. licensing, documentation, training, software tool support);
- 4) **Co-ordinating with other bodies** that deal with biomedical semantic interoperability (e.g. in the area of EHRs, pharmaceutical discovery pipelines (e.g. <http://www.pistoiaalliance.org>), and clinical biosimulation (e.g. <http://www.vph-noe.eu>);
- 5) Conducting **independent benchmarking** to compare vocabulary products for similar domains;
- 6) Supporting efforts towards the (i) **alignment of extant O/Ts** and (ii) **convergence** on a single high-quality O/T resource for each domain;
- 7) Providing **scientific advice** to specific O/T development consortia, and establish objective criteria (e.g. tangible interoperability benefits) for the product **quality certification** of a vocabulary for a specific domain indication;
- 8) Establishing an **open semantic interoperability standard development process** that defines the minimum ontological properties required to represent a particular biomedical domain and involve appropriate scientific, software engineering, clinical and industry stakeholders;
- 9) Studying the **economics of O/T demand and supply**, and the key causes that interfere with (i) O/T adoption or (ii) the realization of the interoperability benefits following adoption.
- 10) Assisting with the **global dissemination, adoption and community engagement** of both standards and vocabularies.
- 11) Supporting the development and use of **structured reporting at source** (e.g. archetypes and clinical data structures for clinical records), as a way to provide a timely and well defined context to O/T terms used during the annotation process;
- 12) Involving the **WHO and national e-health agencies** in establishing a communal understanding of the semantic interoperability problem and promoting community-wide standardization of thoroughly tested and technically coherent methods.

BdB:EBI: 26/10/2009