

## **OBI jamborees: community-driven development of the Ontology for Biomedical Investigations**

This application seeks support for core networking activities designed to facilitate the coordinated development of stable resources, the Ontology for Biomedical Investigations (OBI) and its community oriented environments ("views"), that will together enable an integrated approach to the collection, depositing, curation, management, sharing, mining and reasoning of heterogeneous data from diverse bio-domains. We propose to support the synergetic activities of international, community-driven standard initiative led by the OBI Consortium and coordinating of UK input into the OBI project by organizing a series of workshops. In full alignment with the objectives of the Strategic Tools and Resources Development Fund - "the development of standards (including coordinating UK input into international standards development activities)", - these activities will bring together communities from around the globe for collaborative purposes in order to develop and promote standards in the biological areas of strategic importance.

OBI is a large collaborative project initiated in 2005. OBI aims to support the detailed description of biomedical investigations. It covers all phases of the investigation process, such as planning, execution and reporting. OBI supports data annotation, repository querying and multi-disciplinary data analysis, thereby tackling the challenges presented by the large scale multi-omic data that now characterize biomedical science. OBI represents a substantial and significant effort to describe biomedical investigations in a consensual, wide-ranging and computationally amenable way. Ontologies such as OBI are essential to rise to the challenges of large-scale biology, safeguard the inheritance of experimental data for the future, and maximize its usefulness in the present.

The OBI project is an open collaborative effort. Currently the OBI Consortium combines 19 international research communities and the number of participants is growing. Many of OBI developers are from UK (in alphabetical order): Dr Helen Parkinson (the EBI), Dawn Field (NERC Centre for Ecology and Hydrology, Oxford), Frank Gibson (Abcam plc, Cambridge), Phillip Lord (School of Computing Science, Newcastle University), Allyson L. Lister (School of Computing Science & Centre for Integrated Systems Biology of Ageing and Nutrition, Institute for Ageing and Health, Newcastle University), James Malone (EMBL-EBI, Wellcome Trust Genome Campus, Hinxton), Helen Parkinson (EMBL-EBI, Wellcome Trust Genome Campus, Hinxton), Norman Morrison (School of Computer Science, University of Manchester & NERC Environmental Bioinformatics Centre, CEH Oxford), Philippe Rocca-Serra (OeRC, University of Oxford), Susanna-Assunta Sansone (OeRC, University of Oxford), Larisa N. Soldatova (Aberystwyth University, Wales).

The BBSRC has already supported a number of UK-based OBI workshops, i.e. within the "Omics Data Standards" project led by Susanna-Assunta Sansone. As the result of intensive five years work, supported by a number of funding bodies from around the globe, including the BBSRC, the OBI Consortium is now ready to make freely available the first full release OBI 1.0.

The OBI project is now moving to a new stage OBI 2.0, where it would face new administrative, maintenance, and sustainability tasks:

- Development of community-oriented OBI-views. OBI is complex, but then so are 'materials and methods' sections as they stand. Biologists need user-friendly environments/interfaces masking this complexity from end users. Environments can set the minimal fields required and restrict each of them to appropriate terms from OBI or other OBO ontologies without requiring the end user to have in-depth understanding of these. Interfaces also can display, as alternative to the OBO names, labels which are accepted in a particular community. Such environments are essential for acceptance of OBI as an integrated standard for bio-domains.

- Promotion of OBI as a unified standard for the area of biology and biomedicine. Engagement with potential users and those who govern them (journals, funders, regulators) and industry.
- Further development of OBI without interruption of stable OBI-based services to the communities. Biological investigations are very complex and OBI still is far from sufficient coverage of all the essential components of such investigations. OBI will need to deal with multiple requests for standard representation of entities from bio-domains efficiently and on regular basis. OBI development is driven by specific examples of experiments from contributors which are then generalized. As such, OBI will continue to evolve as new examples are introduced.
- The OBI Consortium needs to ease the submission of new terms and suggestions for corrections from the communities to OBI.
- Further investigation of the application of OBI descriptions to databases, for example the Chemical Effects in Biological Systems (CEBS), in order to support queries over such data.
- Regular (most likely monthly) release of OBI. The OBI Consortium needs to decide upon release policy: what metadata for released terms to provide, what scripts to use. For each release, all OBI branches which are often being developed in parallel, have to be merged, the merged OBI has to be checked for logical consistency and completeness.
- Engagement with software developers. While Protégé has worked well as an editing tool, it requires craft knowledge for use with an ontology of this complexity. In particular, integration with versioning systems, essential for distributed development, is currently limited. Similarly, OWL has enabled computational reasoning which is valuable for both use and development of the ontology. However, current reasoning tools do not provide a transparent explanation and debugging facility; likewise, performance is unpredictable and can change significantly between versions of OBI. Reasoning over large data sets is problematic. The community continues to improve these tools and OBI has benefited from communication with their developers. OBI contributes to the identification of limitations of tools, prepares recommendations and requirements.
- The OBI project will need to be aligned with other, including newly emerging, related projects. There is a number of ongoing synergetic activities. For example, the eagle-i project aims to build a prototype of a national research resource discovery network – one that will help biomedical scientists search for and find previously invisible, but highly valuable, resources. The eagle-i Consortium, made up of nine member institutions, is collaborating with the OBO Consortium in order to provide standardized representation of devices and its functions. [suggestion from Philippe: put collaboration with IEDB, BII, CEBS]
- Contribution to the OBO Foundry. OBI became one of the driving projects in the establishment of the OBO Foundry. OBI will continue the work on the integration of OBO ontologies into a single framework. OBI itself, when mature enough, will be submitted to the OBO as a foundry ontology.

Our experience shows that without regular face-to-face meetings, the development, administration and promotion of OBI would not be possible. Most of work on OBI is done on voluntary basis because the OBI developers realize that the lack of a stable integrated standard for bio-domains is an unacceptable state of affairs for collection, depositing, sharing and integration of research data. The collaborative nature requires frequent face-to-face meetings to create the necessary conditions for the building of consensus.

This proposal seeks funds to manage the process of consensus building through organizing 6 bi-annual workshops in Aberystwyth, Birmingham, Oxford,... [depends on letters of support]. The workshops programme will be in accordance with the tasks of the OBI2-stage listed above. The number of delegates will be limited to 20-22.