A Cross-Species Anatomy Resource XSPAN: Models of Anatomy in the Gene Ontology

J.S. Aitken
School of Informatics, University of Edinburgh,
Artificial Intelligence Applications Institute,
Appleton Tower, Crichton St. Edinburgh EH8 9LE

Abstract

This paper presents the background to the XSPAN project, the approach adopted and the results to date.

The XSPAN project is building on the Gene Ontology initiative to provide tools and data to support the identification of tissue homologies across model species. While the Gene Ontology (GO) provides a vocabulary to describe genes and gene products in terms of their molecular function, process and cellular location, it is not intended to include species-specific concepts such as anatomies and homologies. However, anatomies for model species including C Elegans, Drosophila, Mouse and Zebrafish are being specified in the GO syntax and are being linked to gene expression data by their respective research communities. These anatomies can be seen as species-specific extensions to GO that similarly enable annotation and reasoning.

The XSPAN project is acquiring knowledge of tissue homologies, which is expressed in terms of links between (identified and labelled) tissues in the anatomies of model species. Initial results of knowledge acquisition and modelling in XSPAN include the use of a Cell Function ontology to further specify the basis of a homology link - where the cell function ontology will itself be contributed to GO. Analysis of the semantics of GO ontologies, with a specific focus on the description of anatomy has revealed several areas where clarification is required as to the meaning and definition of the relationships. Clarification of semantics is not an end in itself, rather it is a requirement if our aim is to permit automated reasoning about the inheritance of properties, for example, cell composition/function and the localisation of gene expression, along specialisation and partonomic links. This form of reasoning is required if functions beyond information retrieval are to be implemented in support of the biologist's task of gene expression data analysis.

The technical contributions of XSPAN include the development of a tool to acquire homology mappings using existing anatomies and GO resources. The resulting mappings will be made available through both database, and graphical user interfaces. We have a proposal for a RDFS encoding of anatomies, as a preliminary step to a Web Ontology Language (OWL) encoding - the coming standard for the Semantic Web.

In summary, we adopt a constructive approach to the re-use and formalisation of existing Bio-Ontology resources. While analysis shows some weaknesses in these semi-formalised knowledge sources, our proposals provide both clarification, and the steps specify newly-contributed knowledge in a more formally tractable manner, through the use of a new domain ontology (of cell function), and encodings in a language with more semantics (RDFS).

Acknowledgments

The author acknowledges the support of Prof Jonathan Bard and Prof Bonnie Webber of the University of Edinburgh, and the BBSRC. The Cell Function Ontology is a collaboration between Jonathan Bard and Michael Ashburner, EBI Cambridge, UK.