Practical Implications of Ontology and Global Standards for Model-Based Data Analysis

Chris Franklin (1), Bernard de Bono (2), Pierre Grenon (2), Sarala Wimalaratne (2), Lourdes Cucurull-Sanchez (3), Oscar Della Pasqua (1)
(1)Clinical Pharmacology Modelling and Simulation, GlaxoSmithKline, Stockley Park West, Uxbridge, Middlesex UB11 1BT, UK
(2)European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK
(3)Computational Biology, GlaxoSmithKline Medicines Research Centre, Stevenage, Hertfordshire , SG1 2NY, UK

Introduction

Reusing existing models and data provides one of the biggest benefits to the pharmaceutical industry both in terms of efficiency and increased understanding of drugs and diseases. In order to enable reuse, however, we need to be able to consistently search for appropriate information, retrieve and apply it across different situations (e.g., studies, populations, indications) and environments (e.g., software, platforms, business units, institutions).

DDMoRe is looking to implement the required technologies and an infrastructure for managing a library of models and data to allow sharing of information across the community. To this end, in particular, consensual representation of relevant information will be develop in the form of ontologies. **Ontologies** are formal organisation of domain knowledge, and in that way enable knowledge sharing between different software applications. They provide machine-understandable meaning to its data (without the need for intermediate human interference).

While ontologies have evolved in systems biology and physiology, enabling a framework for relating semantic information and interoperability between models and software applications, little attention has been paid to its importance in model-based drug development.

Our aim, here, is to illustrate the potential impact of ontologies in support of model-based drug development using selected real-life examples.

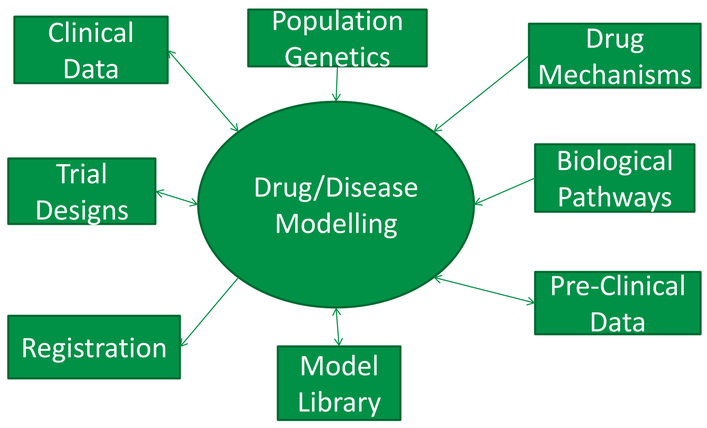

Figure 1 – The modelling environment

Ontologies in Drug Development

Drug development encompasses various domains of knowledge (Figure 1). There already exists a number of ontologies covering some of these domains, in particular:

- Systems Biology – Examples include Systems Biology Ontology and Foundational Model of Anatomy.
- Genetics – Examples include the Gene Ontology Cell Components and Gene Ontology Biological Process.
- Clinical Data – Examples include SNOMED and NCI (National Cancer Institute) CDISC components.

It is consistent with best practice to reuse, whenever possible, existing ontologies. As illustrated above, drug development covers a large range of fields which impose the need for extension and/or integration of available ontologies. Indeed, to be effective, these ontologies must be interconnected and articulated.

The result can be seen as a consistent *encyclopaedia* of all the relevant knowledge covering the domains in a way that is understandable by appropriate software. In the context of DDMoRe, such an *encyclopaedia* will contain all domains mentioned in Figure 1.

Individual usage of ontologies provides contextual classification of the information – including the properties, relationships and instances (e.g. Individual or population).

Methods and Results

Three different example models were selected:

- PBPK Rat model using SimBiology/MatLab
- One Compartment PK model using NONMEM
- PKPD model for NONMEM

To ensure interoperability, these models were inspected to identify the mechanism in which mark-up could be included as well as which elements should be marked up.

The mark-up of models is the concrete way of recording annotations that make explicit the knowledge content in a model. It is such records that power the search capabilities of a library of models (and data) and consequently enable reuse.

To annotate a model means encoding information about it and its elements (e.g., defining the anatomical extent of a compartment) in a machine processable way. A basic mechanism of annotation was utilised, namely the Resource Description Framework (RDF) notation.

Descriptions were added as comments to NONMEM code and data, and as "NOTE" elements within the MATLAB code for the PBPK model. See Figures 2 and 3 for example annotations.

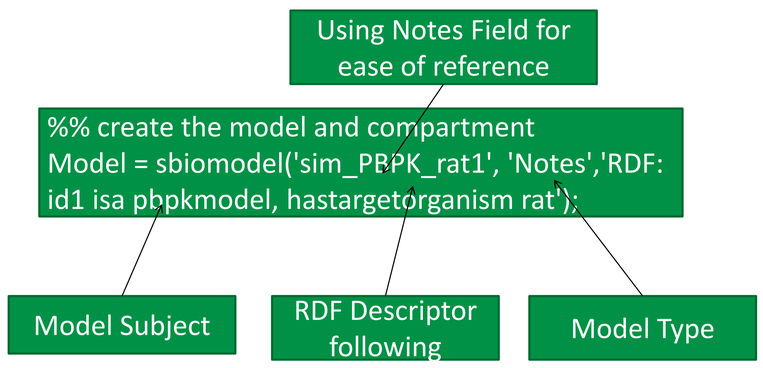

Figure 2 – Example mark-up of a PBPK model

DDMoRe have identified many existing domains for data definitions, see Figure 4. Clinical Data forms an important part of modelling, and the inclusion of CDISC based ontologies is essential. On the other hand some domain ontologies are incomplete for the modelling tasks and will be expanded to include additional information (e.g. statistical methods).

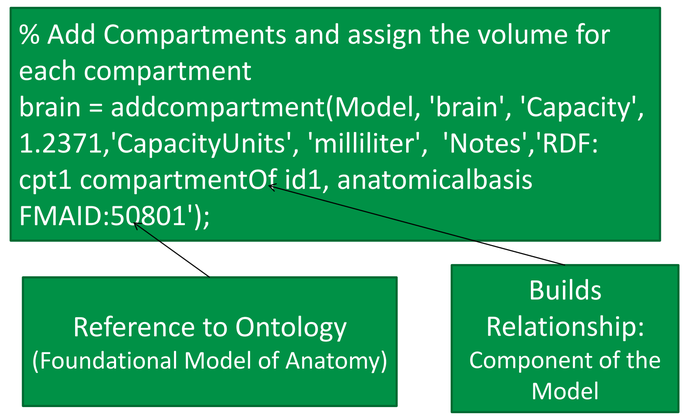

Figure 3 – Example mark-up of a model element with ontology

Using Ontologies

Applying ontologies to models will formalise the use of standards across the R&D community, with clear benefits for model libraries and re-use of models.

Similar standardisation issues have been tackled in other areas of clinical research (e.g. safety). SNOMED allows for common definitions of adverse events and medications during regulatory submissions. In systems biology, the "Open Biological and Biomedical Ontologies" provides an *encyclopaedia* of available ontologies that can be used to search knowledge bases.

Some important questions that could be addressed by applying ontologies within the modelling space include:

- Is there a PBPK, PK and/or PKPD model for drug X?
- Is there a disease model for disease Y?
- Do any models for drug X include Blood Pressure?
- Is there a model of a specific target mechanism embedded within a disease model?

Common specification will also ensure that the underlying assumptions in each model are effectively shared between modellers regardless of their specific area of expertise and/or application.

Conclusion

Models with appropriate ontological annotations have a potential to greatly facilitate the adoption of libraries.

The use of systems biologic terms will allow mechanistic models to be included into earlier stages of the drug-development process, whilst CDISC annotation will be required to address late stage drug development requirements.

A new "modelling and simulation" domain will be required to hold specific terminology relating to the types of pharmacometric models.

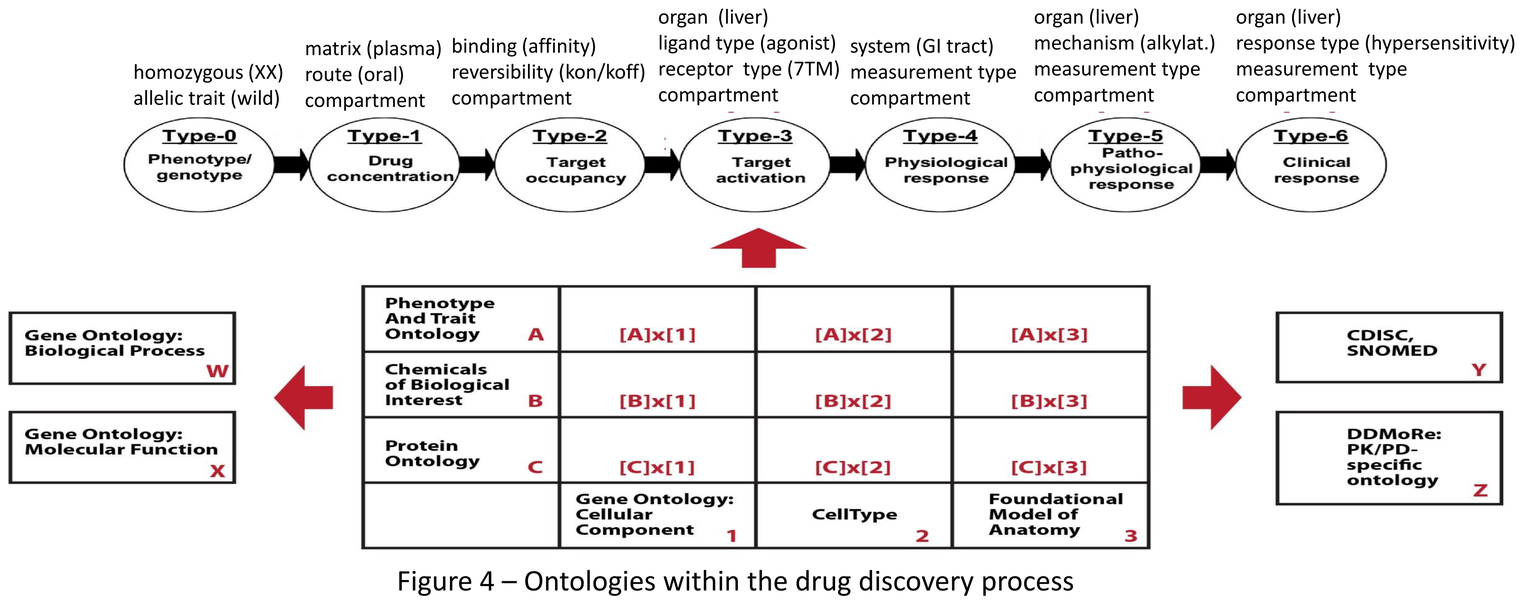

Figure 4 – Ontologies within the drug discovery process

Acknowledgement: The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115156, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. The DDMoRe project is also financially supported by contributions from Academic and SME partners.