

FAIRDOM approach for semantic interoperability of systems biology data and models

Olga Krebs¹, Katy Wolstencroft³, Natalie Stanford², Norman Morrison², Martin Golebiewski¹,
Stuart Owen², Quyen Nguyen¹, Jacky Snoep², Wolfgang Mueller¹, and Carole Goble².

¹ Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

² School of Computer Science, University of Manchester, UK

³ Leiden Institute of Advanced Computer Science, Leiden, NL

ABSTRACT

The ability to collect and interlink heterogeneous data and model collections is essential in Systems Biology. Effective data exchange and comparison requires sufficient data annotation. This is particularly apparent in Systems Biology, where data heterogeneity means that multiple community metadata standards are required for the annotation of a whole investigation, including data, models and protocols.

Here we describe FAIRDOM (<http://fair-dom.org/>) strategy in the context of semantic data management in the openSEEK, a web-based resource for sharing and exchanging Systems Biology data and models.

1 INTRODUCTION

Data integration is an essential part of Systems Biology. Scientists need to combine different sources of information in order to model biological systems, and relate those models to available experimental data for validation. Currently, only a small fraction of the data and models produced during Systems Biology investigations are deposited for reuse by the community, and only a smaller fraction of that data is standards compliant, semantic content.

By embedding semantic technologies into familiar data management tools, the SEEK [1] enables semantic annotation of new data and the generation and querying of linked-data compliant datasets, whilst hiding the complexities of ontologies and metadata from its users. The SEEK is based on the ISA infrastructure (Investigations, Studies and Assays), a standard format for describing how individual experiments (assays) are aggregated into wider studies and investigations [2]. This poster will present the semantic data integration in SEEK, and how it supports the whole life cycle of data collection, annotation, sharing and reuse of Systems Biology data and resources.

2 THE JERM ONTOLOGY AND JERM TEMPLATES

The JERM Ontology is an application ontology designed to describe the relationships between items in SEEK (for ex-

ample, data, models, experiment descriptions, samples, protocols, standard operating procedures and publications); and to enable these relationships to be expressed with formal semantics. It is based on the idea of the Minimal Information Models (<https://www.biosharing.org>), which have been collected under the umbrella of MIBBI (Minimum Information for Biological and Biomedical Investigations). The JERM takes the specification one step further, expressing the minimum information model as an OWL ontology (<http://bioportal.bioontology.org/ontologies/JERM>).

The majority of laboratory scientists use spreadsheets for the daily management and manipulation of data, so the RightField semantic spreadsheet application [3] (also developed during this work) is used to embed semantic annotation into the data. RightField-enabled spreadsheets allow the collection of semantic information by stealth.

By embedding the JERM metadata model in a spreadsheet format, and enabling the use of JERM (and other) vocabulary terms for annotation, the process of standardized semantic data collection can become part of the existing data management activities in the laboratory. JERM spreadsheet templates have been developed for a wide range of experimental data types.

REFERENCES

1. Wolstencroft, K., Owen, S., du Preez, F., Krebs, O., Mueller, W., Goble, C. and Snoep, J.L. (2011) The SEEK: a platform for sharing data and models in Systems Biology. *Methods Enzymol*, 500, 629-655.
2. Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., Hofmann, O. et al. (2010) *ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level*. *Bioinformatics*, 26, 2354-2356.
3. Wolstencroft, K., Owen, S., Horridge, M., Krebs, O., Mueller, W., Snoep, J.L., du Preez, F. and Goble, C. (2011) *RightField: embedding ontology annotation in spreadsheets*. *Bioinformatics*, 27, 2021-2022

* To whom correspondence should be addressed: olga.krebs@h-its.org