

Share FAIR – Data management and standards for Systems Medicine

Martin Golebiewski¹, Olga Krebs¹, Hadas Leonov¹, Stuart Owen², Maja Rey¹, Natalie Stanford², Andreas Weidemann¹, Alan Williams², Ulrike Wittig¹, Katy Wolstencroft³, Jacky Snoep⁴, Wolfgang Müller¹ and Carole Goble²

Heidelberg Institute for Theoretical Studies, Heidelberg, Germany¹, The University of Manchester, Manchester, United Kingdom², Leiden University, Leiden, the Netherlands³, Stellenbosch University, Stellenbosch, South Africa⁴

FAIRDOM develops and offers integrated data management support for research in the fields of systems biology and systems medicine within and across research consortia, like the German research network ‘Systems Medicine of the Liver’ (LiSyM: <http://www.lisym.org>), or European research networks like ERASysAPP and NMTrypI (New Medicines for Trypanosomatidic Infections). Parts of these solutions are also applied to projects with a local focus as the Synthetic Biology Centres at Manchester (SynBioChem) and Edinburgh (SynthSys).

Our data management concept aims at bundling, storing and integrating research assets like data, models and description of processes and biological samples in a Findable, Accessible, Interoperable and Reusable (FAIR) manner (<http://fair-dom.org>) and consists of 4 major pillars:

- 1) Infrastructure backbone: The SEEK software as registry for data, models, samples, processes, publications and presentations, and as yellow pages for projects, people and events. SEEK is either implemented as data management platform that is maintained by the research project itself (e.g. LiSyM SEEK: <http://seek.lisym.org>) or as hub service maintained by us and spanning different consortia (FAIRDOMhub: <https://www.fairdomhub.org>).
- 2) Standardized data description: Data spreadsheet templates and tailored use of controlled vocabularies and ontologies to describe data and metadata
- 3) Modelling support: Seamless handling and simulation of models by integrated modelling platforms (e.g. JWS-Online or SYCAMORE)
- 4) Social support: Facilitators (PALs) in the projects for gathering requirements and dissemination

SEEK specifically supports the interaction between modelling and experimentation. Datasets can be associated with models and/or workflows or biological samples, and model simulations can be compared with experimental data.

For their integration into models, standardization of these data is crucial and also comprises the standardized description of applied methods, biological material and workflows for data processing, analysis, exchange and integration (e.g. into computational models), as well as the setup, handling and simulation of the resulting models. Hence, standards for formatting and describing experimental data, applied workflows and computer models have become important, especially for data integration across the biological scales for multiscale approaches. To this end many grassroots standards for data, models and their metadata have been defined by the scientific communities and are driven by standardization initiatives such as the Computational Modeling in Biology Network (COMBINE). With our activities we also aim at enhancing the harmonization of these standards. This is achieved by building a bridge between

stakeholders and developing the means and channels for transferring information about standards between them, such as the NormSys registry for modelling standards (<http://normsys.h-its.org>).