

FAIRDOMHub: a repository and collaboration environment for sharing systems biology research

Katherine Wolstencroft¹, Olga Krebs², Jacky L. Snoep^{3,4}, Natalie J. Stanford⁴, Finn Bacall⁴, Martin Golebiewski², Rostyk Kuzyakiv⁵, Quyen Nguyen², Stuart Owen⁴, Stian Soiland-Reyes⁴, Jakub Straszewski⁶, David D. van Niekerk³, Alan R. Williams⁴, Lars Malmström⁵, Bernd Rinn⁶, Wolfgang Müller² and Carole Goble^{4,*}

¹Leiden Institute of Advanced Computer Science, Leiden University, Leiden, 2333 CA, Netherlands, ²Heidelberg Institute for Theoretical Studies gGmbH, Schloss-Wolfsbrunnengasse 35, 69118, Heidelberg, Germany, ³Department of Biochemistry, University of Stellenbosch, Private Bag X1 7602 Matieland, South Africa, ⁴School of Computer Science, The University of Manchester, Kilburn Building, Oxford Road, Manchester, M13 9PL, UK, ⁵University of Zurich, Winterthurerstrasse 190, Y12F64, 8057, Zurich, Switzerland and ⁶ETH Zurich, Weinbergstrasse 11, 8092 Zurich, Switzerland

Received August 14, 2016; Revised September 21, 2016; Editorial Decision October 07, 2016; Accepted: October 31, 2016

ABSTRACT

The FAIRDOMHub is a repository for publishing FAIR (Findable, Accessible, Interoperable and Reusable) Data, Operating procedures and Models (<https://fairdomhub.org/>) for the Systems Biology community. It is a web-accessible repository for storing and sharing systems biology research assets. It enables researchers to organize, share and publish data, models and protocols, interlink them in the context of the systems biology investigations that produced them, and to interrogate them via API interfaces. By using the FAIRDOMHub, researchers can achieve more effective exchange with geographically distributed collaborators during projects, ensure results are sustained and preserved and generate reproducible publications that adhere to the FAIR guiding principles of data stewardship.

INTRODUCTION

The FAIRDOMHub is a repository for storing and sharing data, models, protocols and publications relating to systems biology research projects. The systems biology approach has an iterative cycle of experimental and modeling analyses. Experimental results inform mathematical model design and refinement, and modeling simulations direct further laboratory experiments. Data are highly heterogeneous and the relationships between multiple different data sets and mathematical models must be clearly maintained. The interlinking of the experimental data, standard operating procedures (SOPs) and models is essential for interpreting and understanding results. There are several well-

established databases for mathematical models or types of experimental data (e.g. omics data and kinetics), but FAIRDOMHub combines data and models and provides services that enable the integration, interlinking and publishing of experimental and modeling results in the context of the overall systems biology experiment, from a project perspective.

The FAIRDOMHub (<https://fairdomhub.org/>) has been developed as a joint action between ERA-Net ERASysAPP (<https://www.erasysapp.eu/>), an EU-wide consortium of applied systems biology, the European Research Infrastructure and Infrastructure for Systems Biology in Europe (ISBE) (<http://project.isbe.eu/>). It builds on and combines the tools, services and expertise of the earlier SysMO-DB and Sybit data management projects that includes the SEEK platform and openBIS. To date, the FAIRDOMHub contains data, SOPs models, publications and other *research assets* from large research programmes and independent projects, including the SysMO consortium, ERASyAPP and Digital Salmon. It contains more than 1260 data files, 170 models and 200 SOPs, relating to 240 publications arising from some 40 projects. The underlying FAIRDOM software platform and related tools are also available for download and independent installation (1,2). Currently 22 independent instances have been deployed.

MATERIALS AND METHODS

The FAIRDOMHub is a public service for users to upload and publish research assets. It serves as a collaboration tool before publication and as a supplementary materials resource to link data, models and other research assets to manuscripts after publication. Projects of all sizes can be registered; from large, multinational consortia to

*To whom correspondence should be addressed: Tel: +44 161 2756195; Fax: +44 161 275 6236; Email: carole.goble@manchester.ac.uk

those undertaken by individual researchers. An example of a published (3) collection of data and models is the ‘Glucose metabolism in *Plasmodium falciparum* trophozoites’ investigation (<https://fairdomhub.org/investigations/56>) that has 3 Studies, 24 Assays, 16 Data files, 19 Models, 13 SOPs and 3 Publications. The Investigation describes the construction, analysis and validation of a model of glycolysis in the malaria parasite *P. falciparum*. The *Model Construction Study* describes the experimental kinetic characterization of each individual reaction. The *Model Validation Study* shows how the model performs and compares to experimental results for the system, and finally, the *Model Analysis Study* points at possible drug targets. This Investigation shows a typical systems biology experiment, with iterations between modelling and experimental data sets. The Investigations, Studies and Assays (ISA) structure on the FAIRDOMHub gives a complete listing of all experimental data and mathematical analyses and makes the model construction and validation transparent and reproducible.

Data sharing policies are upheld through assigned roles that consortia and projects manage themselves. Researchers remain in control of their own assets and can decide who to share them with, and when to publish them. To support data citation and credit each asset is assigned a persistent URL and retains a link to the person and project it came from (e.g. Model 138 is associated with the SARCHI project and two individual researchers from the project – <https://fairdomhub.org/models/138>). For publication, digital object identifiers (DOI) can be minted for individual assets or for a collection that forms part of an investigation (e.g. DOI: 10.15490/seek.1.investigation.56, which is referenced in (3)). This means that content can be directly cited by researchers. Assets relating to projects may also be stored in specialist public repositories such as SABIO-RK (4) for biochemical reactions and their kinetics and BioModels for models (5). Regardless of where they physically reside FAIRDOMHub organizes these assets in one place by linking out to other repositories. It thus aggregates content spread over distributed stores while retaining the context of the investigation; reusing public repositories and securely linking to project stores where data are too large or too sensitive to upload.

Standards

The FAIRDOM platform provides tools and resources to help standards-compliance of uploaded assets to support, for example, the simulation of models and rich querying across the content. The Just Enough Results Model (JERM) is a Minimum Information Model created to describe the interrelations between assets in the FAIRDOMHub and the metadata fields required to describe them. It builds on, and combines, existing life science Minimum Information models (e.g. MIAPE), which aim to capture the least amount of information needed to understand and interpret an experiment, and the ISA structure that allows the aggregation of individual assays into related studies and investigations. JERM templates are defined and shared in spreadsheet format for different types of experimental data using the RightField annotation tool (6) (e.g. https://fairdomhub.org/data_files/1214). A collec-

tion of templates, developed in collaboration with users, are freely available through the Help Documents section. Although all model types are supported, the recommended standard for describing models is MIRIAM-compliant Systems Biology Markup Language (SBML) (7). SBML model components can be matched with data items from uploaded files to find new sources. Model versions can be compared using the BiVes (<https://sems.uni-rostock.de/projects/bives/>) plugin. SBML models are simulated through an integrated JWS Online plugin (8) (e.g. click the ‘Simulate Model on JWS’ button for <https://fairdomhub.org/models/138>). Simulation instructions can be stored in Simulation Experiment Description Markup Language (SED-ML) format (9) as part of a COMBINE Archive (10) that can be downloaded directly from the JWS simulator and is linked to the model analysis through JWS Online.

Access, browsing and searching

Access is through the web browser, or programmatically via the RESTful interface in XML or in Resource Description Framework. External resources can be searched via the FAIRDOMHub interface through available RESTful Application Programming Interfaces (APIs). For example, the BioModels database (5) can be searched together with the models in FAIRDOMHub, allowing users to place results in a broader context. The main web user interface allows users to browse the *Yellow Pages* (registered programmes, projects, institutions and people), the *Experiments* (Investigations, Studies and Assays), the *Research Assets* (Data, Models, Standard Operating Procedures, BioSamples, Organisms and Publications) and *Activities* (presentations and events). Rich browsing enables navigation between and across these categories and a graphical representation shows how individual assets are related to Assays, Studies and Investigations. The interface is designed to be interactive with the user navigating through the graph. Figure 1 shows an *Investigation*, one of its constituent Studies, Assays in that study, and a selection of associated research assets. For each uploaded asset, its version, license and access activity are recorded. If assets are published they are open to users without an account. Others are restricted through a comprehensive permissions system that controls visibility to projects, subgroups and even individuals. In some cases, metadata descriptions are visible, but the content is only available after contact with the owners, via the *Request Data File* link. Registered users can subscribe to specific assets to be notified of changes. Content is indexed on upload and available through a Solr search. Metadata from assets uploaded in JERM-compliant formats is extracted and stored as semantic annotations to assist navigation and search. To compare models and identify potential overlaps, semantic annotations are used to link reaction species to their exact definitions and structures in public databases such as BRENDA, SABIO-RK and ChEBI.

Curation and review

The FAIRDOMHub is a public repository with an open submissions policy. New registrations from consortia are

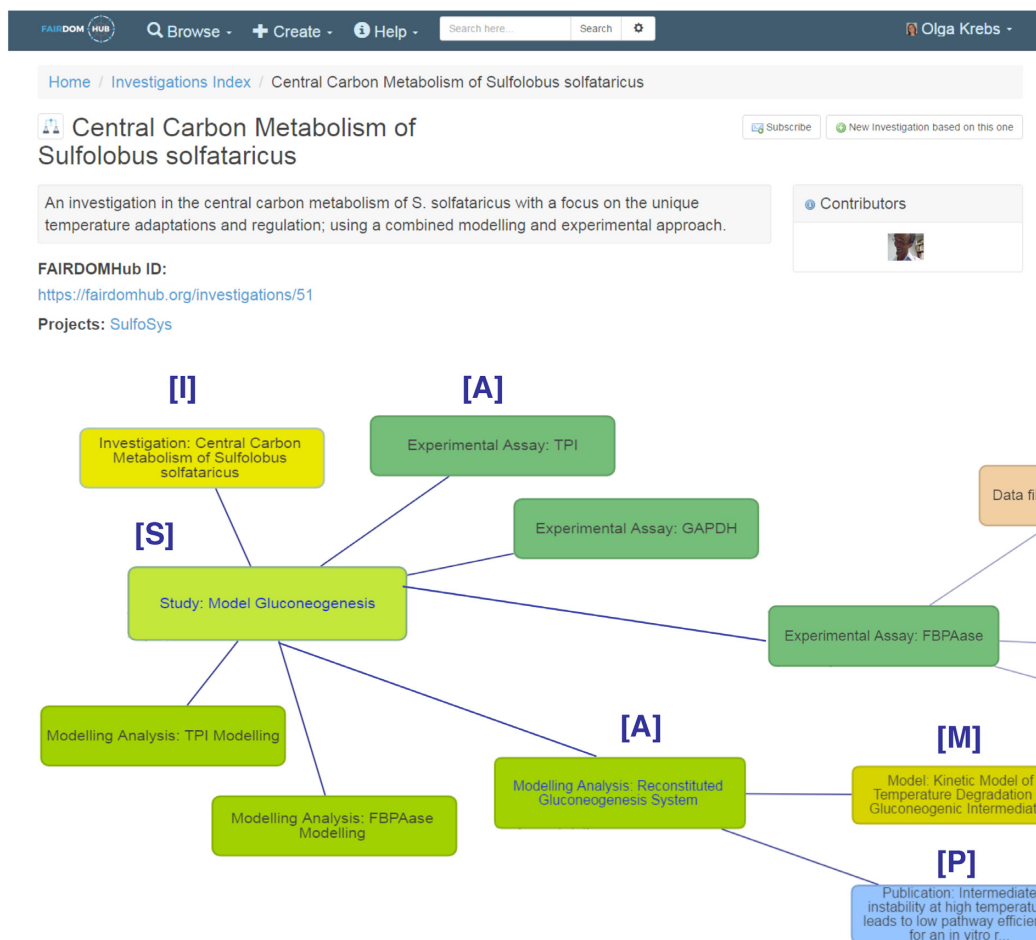


Figure 1. The FAIRDOMHub interface showing an Investigation into the Central Carbon Metabolism of *Sulfolobus solfataricus* (12) (<https://fairdomhub.org/investigations/51>) and the relationships between different assays in one of the constituent studies. The graphical view shows Investigations (I), Studies (S), Assays (A), Data (D), Models (M), Operations (or Standard Operating Procedures-O), and Publications (P). Assays are displayed in two groups; top right shows experimental assays, and bottom left shows modeling assays (analyses). This is a simpler investigation than that of (3) chosen for clarity on the printed page. In practice the interface is designed to be interactive so that users navigate through the graph. Our forthcoming folder-based view will provide an alternative presentation.

activated by a FAIRDOM administrator and then researchers administer their own programmes and projects, assigning membership and project roles. FAIRDOMHub provides curated, standards-compliant templates for data entry and standard recommended formats for models. The majority of assets are shared publicly only when research papers have been accepted. Potential re-users can therefore assume that the data sets and/or models have been peer-reviewed, but not necessarily that their formats and metadata contents have been curated. The FAIRDOM project offers curation services to consortia with direct FAIRDOM project support, including consultations for data structure and formats and model standards compliance. Technical curation of mathematical models published in a number of journals (i.e. *FEBSJ*, *Microbiology*, *Metabolomics*, *IET Systems Biology*) is also provided via the link to JWS Online. Secure access for pre-publication peer review is supported.

Publishing

FAIRDOMHub aims to help researchers share systems biology results along with relevant data, models and protocols to promote reuse and support reproducible publication. For example, support for the SED-ML format, combined with links to the model and publication in the ISA structure, generates a reproducible record of the model simulation events for figures in published articles. Sharing and publishing of individual and collections of assets is governed by assigned credentials; for example sharing permissions can be changed by anyone granted *manage* privileges by the original uploader. 'One-click publishing' enables an asset and selected associated assets to be bundled into a 'snapshot' and assigned a DOI. Snapshot collections can be whole *Investigations* or smaller collections of *studies* and *assays*. The snapshot DOI serves as a supplementary material link in journals. Snapshots can also be packaged into zip files and exported. The snapshots are a form of Research Object (11); digital collections of resources packaged together with a citable identifier and enriched by metadata

describing the collection, and relationships between components. For interoperability the Research Object and the COMBINE Archive specifications have been aligned.

Versioning

For each uploaded asset, its version is recorded; as can relationships indicating if it was derived from, or influenced by, another asset. The version number is displayed with each entry and a comments field allows data uploaders to describe what has changed between versions. The snapshot DOI for publication fixes the version of the assets to be those reported in the paper. Of course assets may continue to be developed, in particular models; users can therefore navigate to subsequent versions through FAIRDOMHub. Plugins support model version visualization and comparison.

DISCUSSION

The FAIR guiding principles of data stewardship (13) promote data sharing and reuse. FAIR adherence is becoming central to several Research Infrastructures within Europe such as ELIXIR (<http://www.elixir-europe.org>) and research funding data mandates increasingly require projects to make their data and models FAIR. Due to the heterogeneous and complex nature of systems biology, researchers in this area need more support to achieve FAIR data compliance. The FAIRDOMHub provides this functionality. It allows systems biologists to share and disseminate their research in a FAIR way that enables experimental data and mathematical modeling results to be presented and interpreted together. It has attracted large national and international consortia and small research groups as users. In addition the FAIRDOM website (<http://www.fair-dom.org>) provides a community ‘knowledge hub’ for data and model management issues, including links to community repositories, standards and initiatives, specialized software, webinars and tutorials and community groups.

ACKNOWLEDGEMENTS

The authors would like to thank Martin Scharm, Dagmar Waltemath, Andrew Millar, project ambassador ‘PALS’ and all others who have supported them in their efforts to develop and deploy the FAIRDOMHub.

FUNDING

Biotechnology and Biological Sciences Research Council [BBG0102181, BB/I004637/1, BB/M013189/1]; Bundesministerium für Bildung und Forschung [0315749,

031A525, 0315781, 031A371]; Klaus Tschira Foundation (KTS); SystemsX; DST/NRF - South Africa [SARCHI 82813 and TTK14051967526]; Netherlands Organisation for Scientific Research [832.14.004]. Funding for open access charge: Biotechnology and Biological Sciences Research Council [BB/M013189/1].

Conflict of interest statement. None declared.

REFERENCES

- Bauch,A., Adamczyk,I., Buczek,P., Elmer,F.J., Enimanev,K., Glyzowski,P., Kohler,M., Pylak,T., Quandt,A., Ramakrishnan,C. *et al.* (2011) openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*, **12**, 468.
- Wolstencroft,K., Owen,S., Krebs,O., Nguyen,Q., Stanford,N.J., Golebiewski,M., Weidemann,A., Bittkowski,M., An,L., Shockley,D. *et al.* (2015) SEEK: a systems biology data and model management platform. *BMC Syst. Biol.*, **9**, 3
- Penkler,G., du Toit,F., Adams,W., Rautenbach,M., Palm,D.C., van Niekerk,D.D. and Snoep,J.L. (2015) Construction and validation of a detailed kinetic model of glycolysis in *Plasmodium falciparum*. *FEBS J.*, **282**, 1481–1511.
- Wittig,U., Kania,R., Golebiewski,M., Rey,M., Shi,L., Jong,L., Algae,E., Weidemann,A., Sauer-Danzwith,H., Mir,S. *et al.* (2012) SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res.*, **40**, D790–D796.
- Li,C., Donizelli,M., Rodriguez,N., Dharuri,H., Endler,L., Chelliah,V., Li,L., He,E., Henry,A., Stefan,M.I. *et al.* (2010) BioModels Database: an enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.*, **4**, 92
- 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.
- Le Novere,N., Finney,A., Hucka,M., Bhalla,U.S., Campagne,F., Collado-Vides,J., Crampin,E.J., Halstead,M., Klipp,E., Mendes,P. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509–1515.
- Olivier,B.G. and Snoep,J.L. (2004). Web-based kinetic modelling using JWS Online. *Bioinformatics*, **20**, 2143–2144.
- Bergmann,F.T., Cooper,J., Le Novere,N., Nickerson,D. and Waltemath,D. (2015) Simulation experiment description markup language (SED-ML) Level 1 Version 2. *J. Integr. Bioinform.*, **12**, 262.
- Bergmann,F.T., Rodriguez,N. and Le Novere,N. (2015) COMBINE archive specification version 1. *J. Integr. Bioinform.*, **12**, 261.
- Belhajjame,K., Zhao,J., Garijo,D., Gamble,M., Hettne,K., Palma,R., Mina,E., Corcho,O., Gómez-Pérez,J., Bechhofer,S. *et al.* (2015) Using a suite of ontologies for preserving workflow-centric research objects. *J. Web Semantics*, **32**, 16–42.
- Kouril,T., Esser,D., Kort,J., Westerhoff,H.V., Siebers,B. and Snoep,J.L. (2013) Intermediate instability at high temperature leads to low pathway efficiency for an in vitro reconstituted system of gluconeogenesis in *Sulfolobus solfataricus*. *FEBS J.*, **280**, 4666–4680.
- Wilkinson,M.D., Dumontier,M., Aalbersberg,I.J., Appleton,G., Axton,M., Baak,A., Blomberg,N., Boiten,J.W., da Silva Santos,L.B., Bourne,P.E. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship scientific data 3. *Sci. Data*, **3**, 160018.