Data Needs Structure: Data and Model Management for Distributed Systems Biology Projects

Martin Golebiewski[1#], Lihua An[1], Iryna Ilkavets[1], Olga Krebs[1], Stuart Owen[2], Quyen Nguyen[1], Natalie Stanford[2,3], Andreas Weidemann[1], Ulrike Wittig[1], Katy Wolstencroft[4], Jacky L. Snoep[2,3,5], Wolfgang Mueller[1] and Carole Goble[2]

[#] martin.golebiewski@h-its.org **http://www.fair-dom.org/**

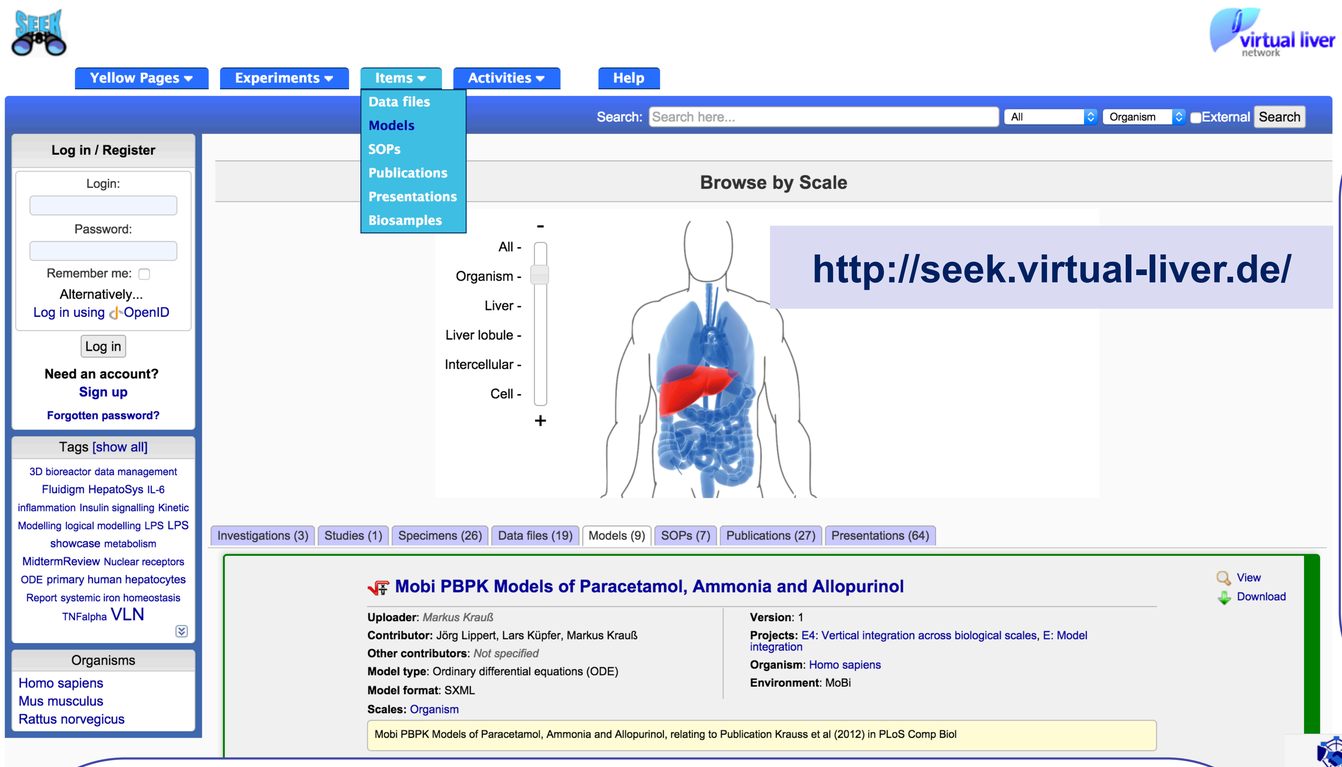
Heidelberg Institute for Theoretical Studies

HITS

FAIRDOM

virtual liver network

ERASysAPP ERA-Net for Applied Systems Biology



Managing Systems Biology Projects

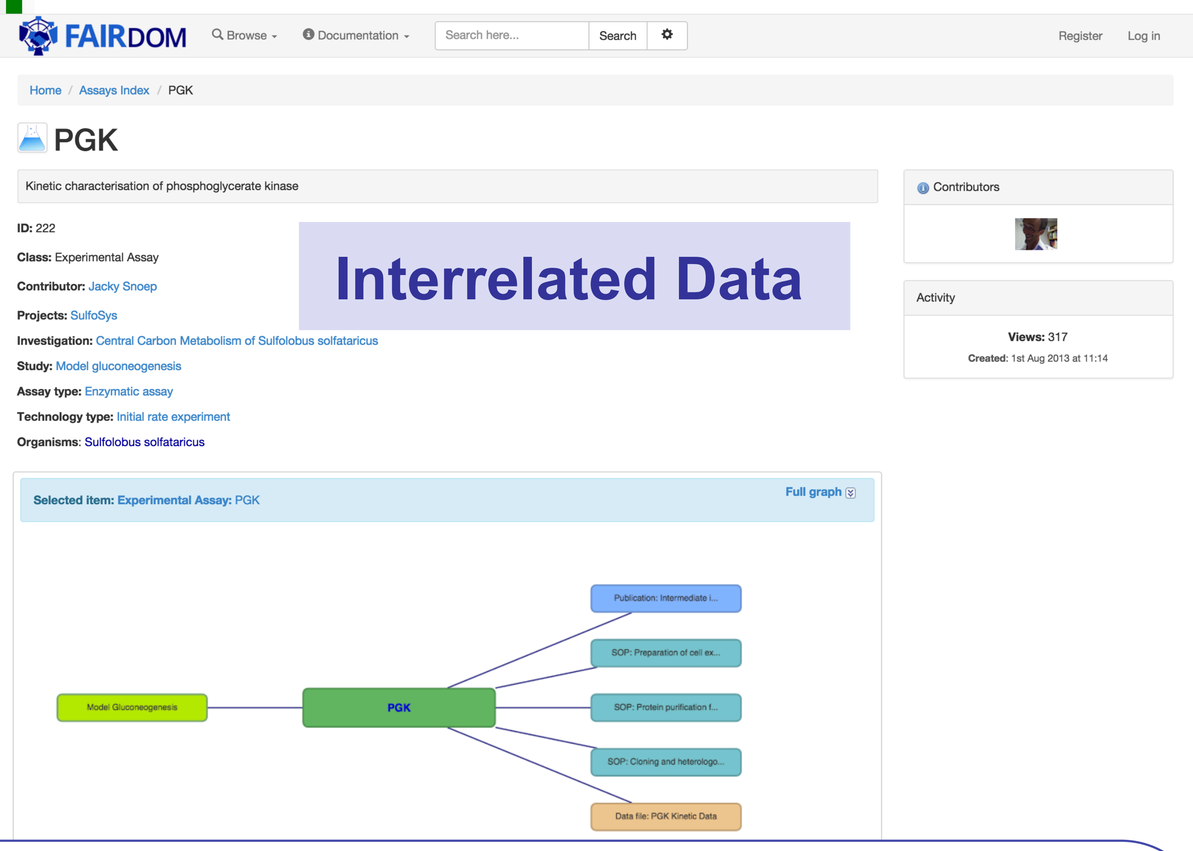
SEEK* is a web-based data management environment for scientists to share and exchange interrelated assets and their corresponding metadata: Data, models, standard operating procedures (SOPs), publications, presentations and biological samples (specimens).

It is also a 'yellow pages' registry for distributed Systems Biology research networks and their elements: Involved scientists, projects, institutions, studies, assays and events. *Wolstencroft *et al.*, BMC Systems Biology, 9:33

SEEKs for Different Communities

Initially developed for **SysMO** (http://www.sysmo.net/) and the German **Virtual Liver Network** (http://www.virtual-liver.de/), SEEK has been adopted as data management hub by more than 10 other consortia, including European research networks like **ERASysAPP** (ERA-Net for Systems Biology Applications) or **NMTrypI** (New Medicines for Trypanosomatidic Infections), as well as Synthetic Biology Centres (**SynBioChem** or **SynthSys**).

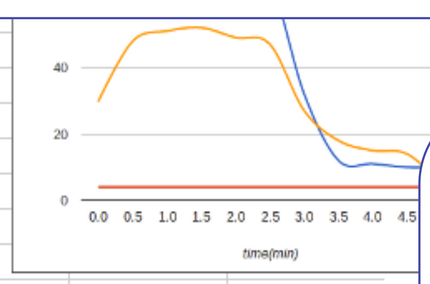
The differences between the consortia drive the development of different interfaces and tools around the data, allowing new user communities to construct a customised SEEK from configurable components, or to install a pre-configured Virtual Machine image.



Interrelated Data

Exploring and plotting data

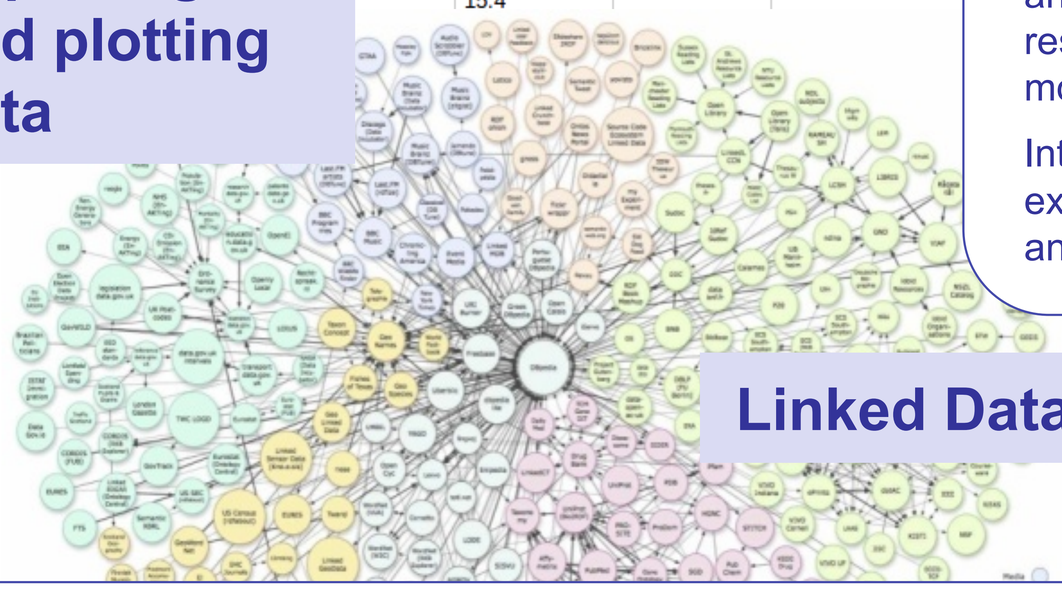
3	0.5	72.4	4.22	48.6
4	1.0	71.2	4.22	51.2
5	1.5	70.0	4.22	52.2
6	2.0	69.2	4.23	49.0
7	2.5	66.0	4.5	47.2
				27.0
				18.0
				15.4



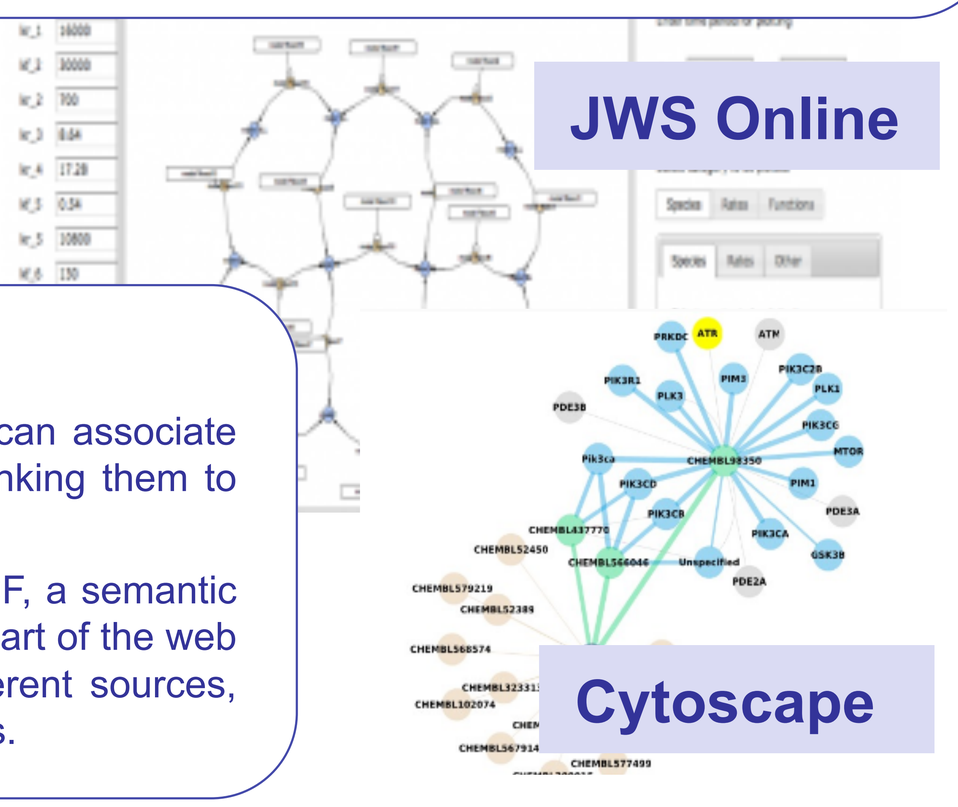
Exploring, Annotating and Simulating Models

Direct simulation of models uploaded to SEEK and standards-compliant annotation, as well as the comparison of simulation and experimental results is possible through the seamless integration of the web-based modelling tools **JWS Online** and **SYCAMORE**.

Integration with other tools, such as **Cytoscape**, or a search gateway to external model databases like **BioModels**, allow further visualisation, analysis and management of models.



Linked Data Cloud



JWS Online

Cytoscape

Exploring, Annotating and Analysing Data

Datasets of any kind can be shared, stored and explored through SEEK. Users can associate corresponding data, plot graphs and add further annotation to data, as well as linking them to relevant models, assays, SOPs, used biological samples, etc.

If standards-compliant data is uploaded to SEEK, it is extracted and stored as RDF, a semantic web format that allows sophisticated querying and enables SEEK data to become part of the web of linked data (the 'semantic web'). This will help in cross-linking data from different sources, even from different SEEK installations and across distinct systems biology initiatives.

Funding:

BBSRC bioscience for the future

Federal Ministry of Education and Research

Affiliations:

[1] Heidelberg Institute for Theoretical Studies (HITS), Germany
[2] School of Computer Science, University of Manchester (UK)
[3] Manchester Institute for Biotechnology, University of Manchester (UK)
[4] Leiden Insitute of Advanced Computer Science, Leiden University (NL)
[5] Department of Biochemistry, University of Stellenbosch (South Africa)

Heidelberg Institute for Theoretical Studies

HITS

The University of Manchester

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot·your knowledge partner