MANCHESTER 1824
The University of Manchester

SEEK for Science:
A Data Management Platform to support Open and Reproducible Science

Professor Carole Goble

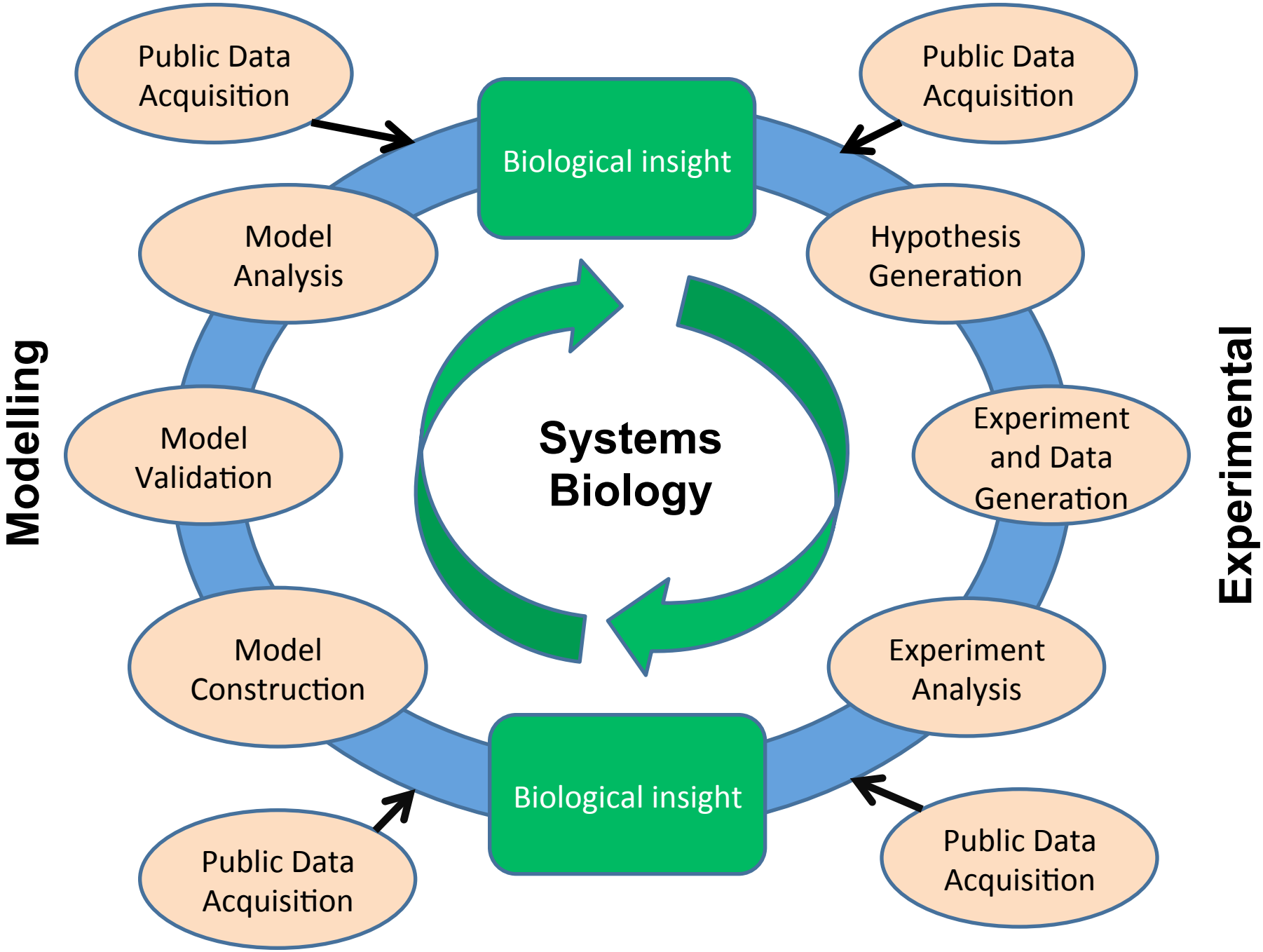
The University of Manchester

UK

BBSRC
bioscience for the future

Federal Ministry of Education and Research

BOSC 2014, 12th July 2014


Public Data Acquisition
Public Data Acquisition
Biological insight
Model Analysis
Hypothesis Generation
Modelling
Experimental
Model Validation
Systems Biology
Experiment and Data Generation
Model Construction
Experiment Analysis
Biological insight
Public Data Acquisition
Public Data Acquisition

Sponsors and Motivation

- EU ERANet programme
- 122 organisations
- 16 multi-inst. consortia
- independent projects in a two-round funding initiative

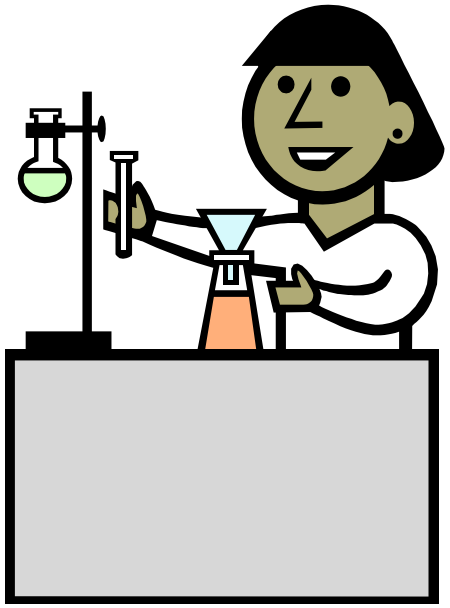
- BMBF “Großprojekt“
- ~45 organisations
- ~70 groups
- multiscale rep of the liver
- multiscale data, models
- imaging data

Funders

- *Preserve* results beyond projects.
- *Organise* & *link* data, models, processes.
- *Exchange* & *search* initiative‘s assets.
- *Share & disseminate* results
- *Improve* standard curation practice.
- *Pool* capacities.
- *Handle home-brewed* solutions with mixed resourcing and no access

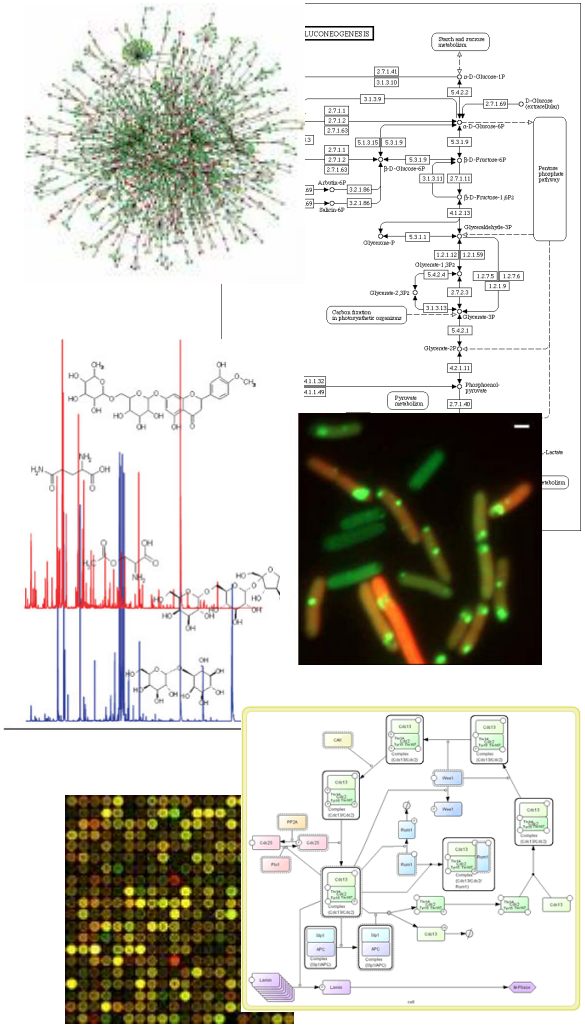
People

- Dynamic distributed groups of modellers and experimentalists.
- Mixed teams of experimentalists and modellers
- Cherished own home-grown and unstable data solutions
  - wikis, CMS, databases, spreadsheets, files)
- Access & visibility control over shared content



Content

- Locally hosted private repositories
- Public archives
- From single-cell to human
- Samples, Specimens, Standard Op Procedures



- Small Data: Reactome…: files, spreadsheets
- Big Data: NGS, Mass Spec…: Specialist repositories, files
- Models: ODE, SBML, Native Matlab, PDE, Multi-scale

- In progress: versioning, track provenance and parameters
- Published: citation, links to publications

The Web-based SEEK Platform

Ruby on RAILS 3.2, BSD,
https://bitbucket.org/seek4science/seek
http://www.seek4science.org

Find, share and exchange **Data**, **Models** and **Processes** within the SysMO Consortium.

Yellow pages | Experiment details | Assets | Activities | Documentation

Data files | Models | SOPs | Publications | Biosamples | All | External | Search

Log in / Register

Login:
Password:
Remember me:
Alternatively...
Log in using OpenID
Log in
Need an account?
Sign up
Forgotten password?

Tags [show all]

Bacillus subtilis Biochemistry Biochemistry and protein analysis Bioinformatics chemostat Computational and theoretical biology Computational Systems Biology Data Management dynamics and control of biological ne... Fermentation Genetic modification Genetics Mathematica Mathematical modelling Matlab Metabolomics Microarray analysis Microbiology Molecular Biology Molecular biology

Home > Models Index > TPI Kinetic Model

TPI Kinetic Model

Download Model

1 item (and an image) are associated with this Model:

Filename: tpi-model.nb
Format: Binary file type
Size: 26.0 KB

Uploader and creators

Attributions
None

Views: 41 **Downloads:** 4
Created: 1st August 2013 at 16:54
Last used: 6th October 2013 at 13:57

Organism: Sulfolobus solfataricus

Model type: Ordinary differential equations

Model format: Mathematica

Execution or visualisation environment: *Not specified*

Model image: *(Click on the image to zoom)*

$$v_{TPI} = \frac{V_{Mf} \cdot \frac{GAP}{K_{M,GAP}} - V_{Mr} \cdot \frac{DHAP}{K_{M,DHAP}}}{(1 + \frac{GAP}{K_{M,GAP}} + \frac{DHAP}{K_{M,DHAP}} + \frac{3PG}{K_{i,3PG}} + \frac{PEP}{K_{i,PEP}})}$$

Description:

Mathematical model for TPI kinetics, GAP and DHAP saturation and inhibition with 3PG and PEP

https://seek.sysmo-db.org/models/114

Cataloguing

Find my peers

Creating and sharing SOPs across projects

Track my specimens

Track different versions of my model

yellow pages, manage SOPs and link them to investigations, studies, assays, specimens and samples

Browse experimental data without downloading them

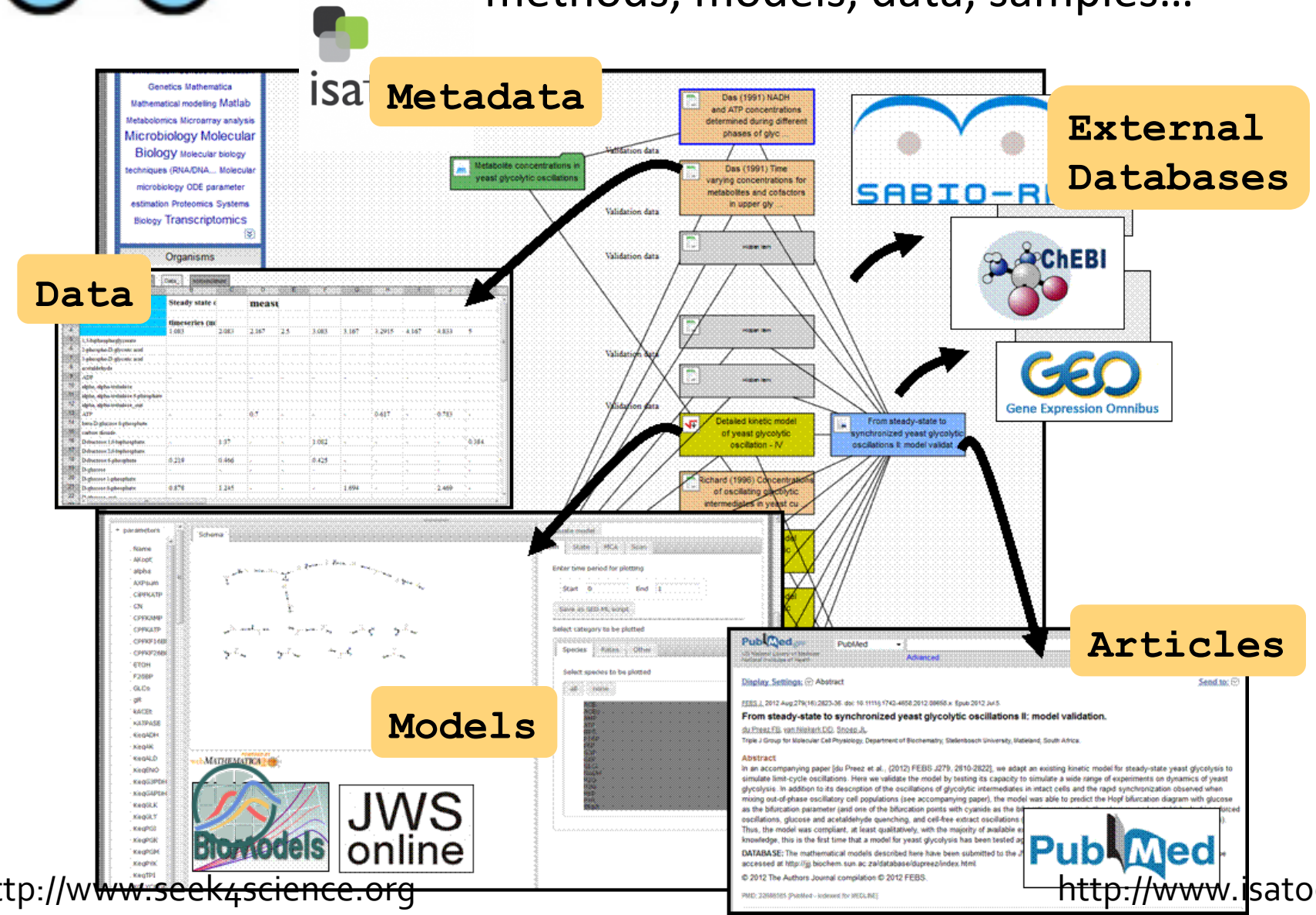
How data, models and SOPs fit together

Which data belong to which publication

Data viewing functionality
ISA: Link Studies to their data, models, SOPs, samples, publications

Aggregated Asset Infrastructure

share and interlinking multi-stewarded, mixed, methods, models, data, samples…



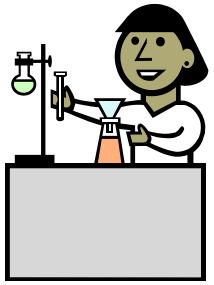
http://www.seek4science.org

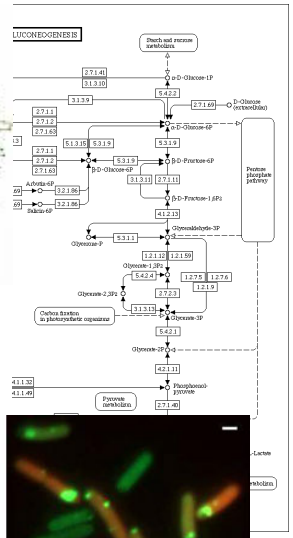
http://www.isatools.org

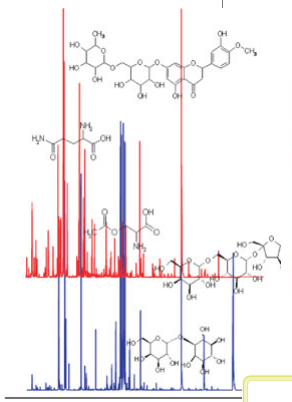
Reproducibility

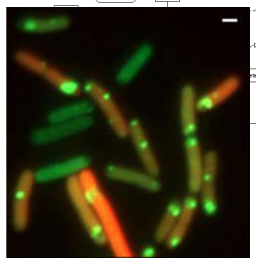
Cross-cataloguing/link, Packaging

- People, experiments procedures, data, models, samples, publications
- Common elements: enzyme rate equations; values for the Km / Vmax (concentration of substrate / maximal velocity of activity speed) of that enzyme









Model simulation

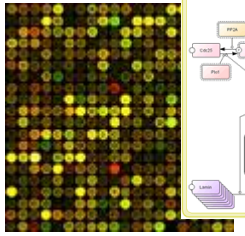
- Integrated simulators & data comparisons
- Differentiate construction, validation & predicted data

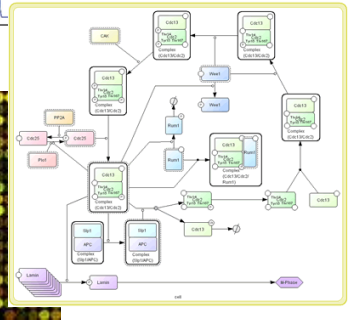
Standards compliance

- SBML, SED-ML, SBGN, MIBBI, ISA, MIRIAM, CellML. COMBINE
- Curation support: tools & services.

Access and Archiving

- Deposition to public repositories
- Gateway/cross-link to public archives & lab repositories
- Public archive for publishers, funders, researchers, trainers.

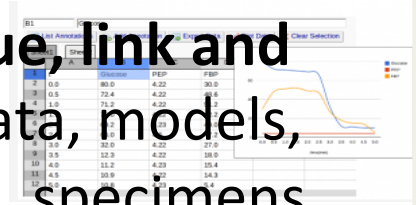




Reproducibility Card

yellow pages of peers projects, experts

catalogue, link and index data, models, samples, specimens, sops, experiments, publications using standards



project mgt, access control reporting, citation governance & policies

simulate models differentiate construction, validation & predicted data

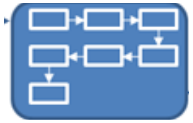
curate & annotate data and models using standards with compliance tools

manage, store and exchange different types and scales of data

integrate local and project tools and data systems

incorporate public data and model repositories & tools deposition

scaled-out collection & **analytics** using third party platforms



Web Interface

REST API

Project DM

Wikis

CMS

Local SEEK

Own DB

JERM Extract, Harvest, Index

SEEK

Yellow Pages People Projects Institutions

Asset Catalogue

Access Privileges

Versions Tags

ISA

Investigation

Study

Assay

Events Presentations

Datafiles Publications

Models SOPs

APIs and Links

Direct Upload

RightField

Datafiles

Models

SOPs

External

BioPortal

PubMed

BioModels

JWS Online

CheBI

SABIO-RK

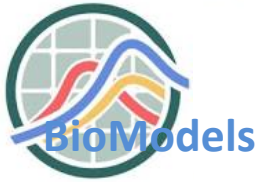
GEO

OpenBIS

Plug-in, Play nice

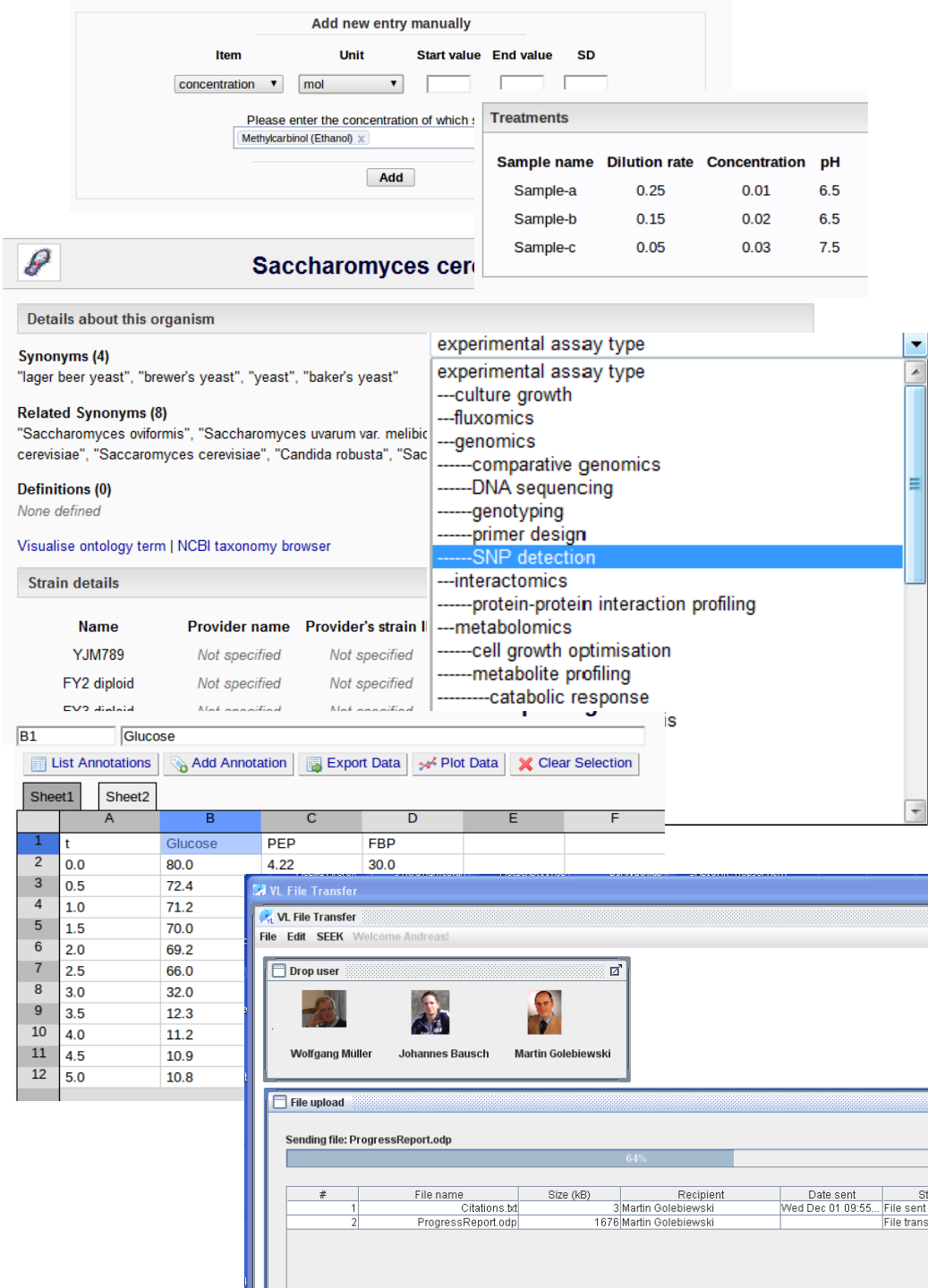
- Gateway plugin framework
  - Tight and loose coupling
  - RAILS plugin or bundled GEM
- Metadata framework
  - JERM and ISA
- Different instances
  - Single query across all model repositories
  - One click deposition

JWS Online
OneStop
DataFuse

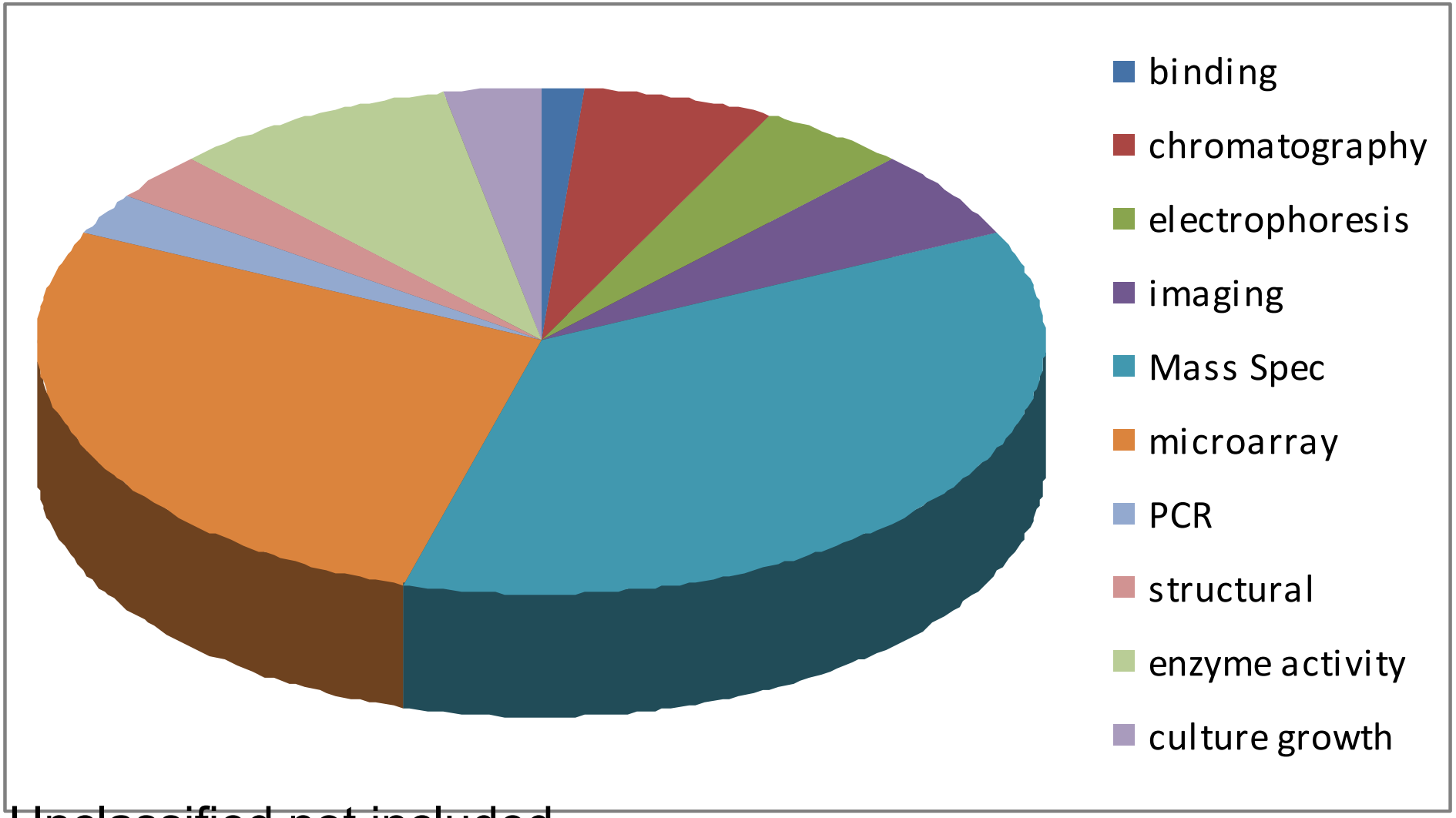


Data….

- Public and new data
- Factors studied
  - Linked -> SABIO-RK and ChEBI
- Samples and Specimens
  - Extends EBI/NCBI BioSamples
- Treatment Extraction
- Tagging with vocabularies
- Spreadsheet-based data-view
- Big Data
  - Upload and by email, Share by trusted link, Link to external repository
- Access
  - DOIs and Temp links for reviews



SysMO projects Assay and Technology Type



Unclassified not included

Models….

Repositories

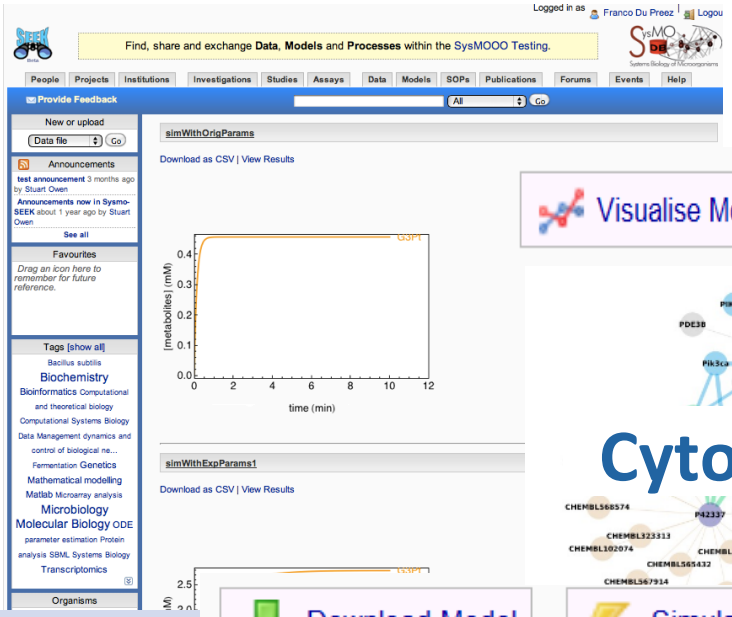
- Biomodels, JWS Online, local SEEK

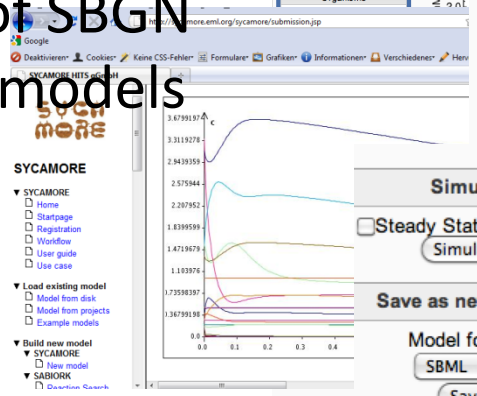
JWS Online Simulator

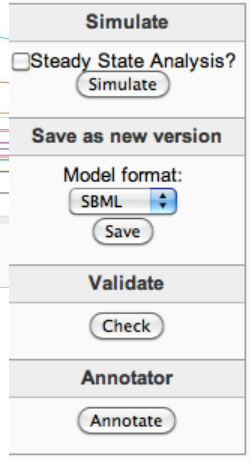
- SBML support
- Auto generation of SBGN schemas for user models
- SED-ML export

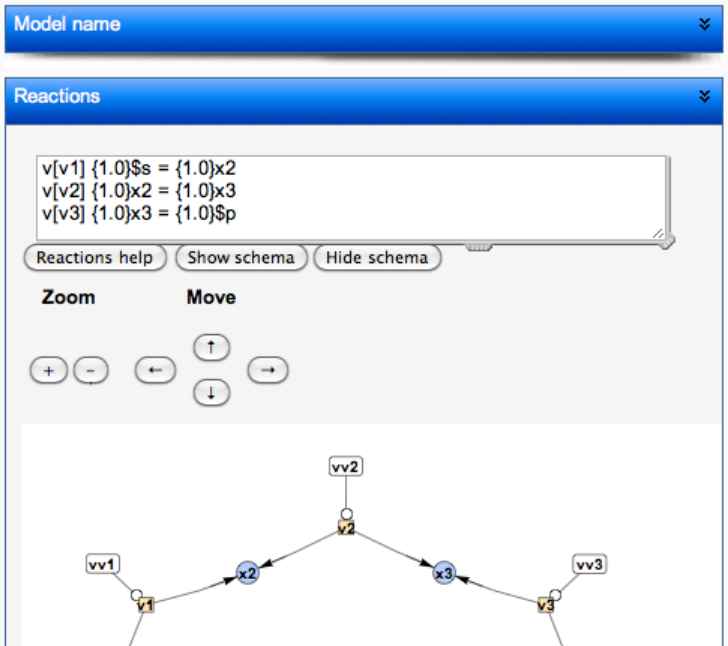
DataFuse

- Link and compare construction and validation data with models
- Run models with parameter values from spreadsheets

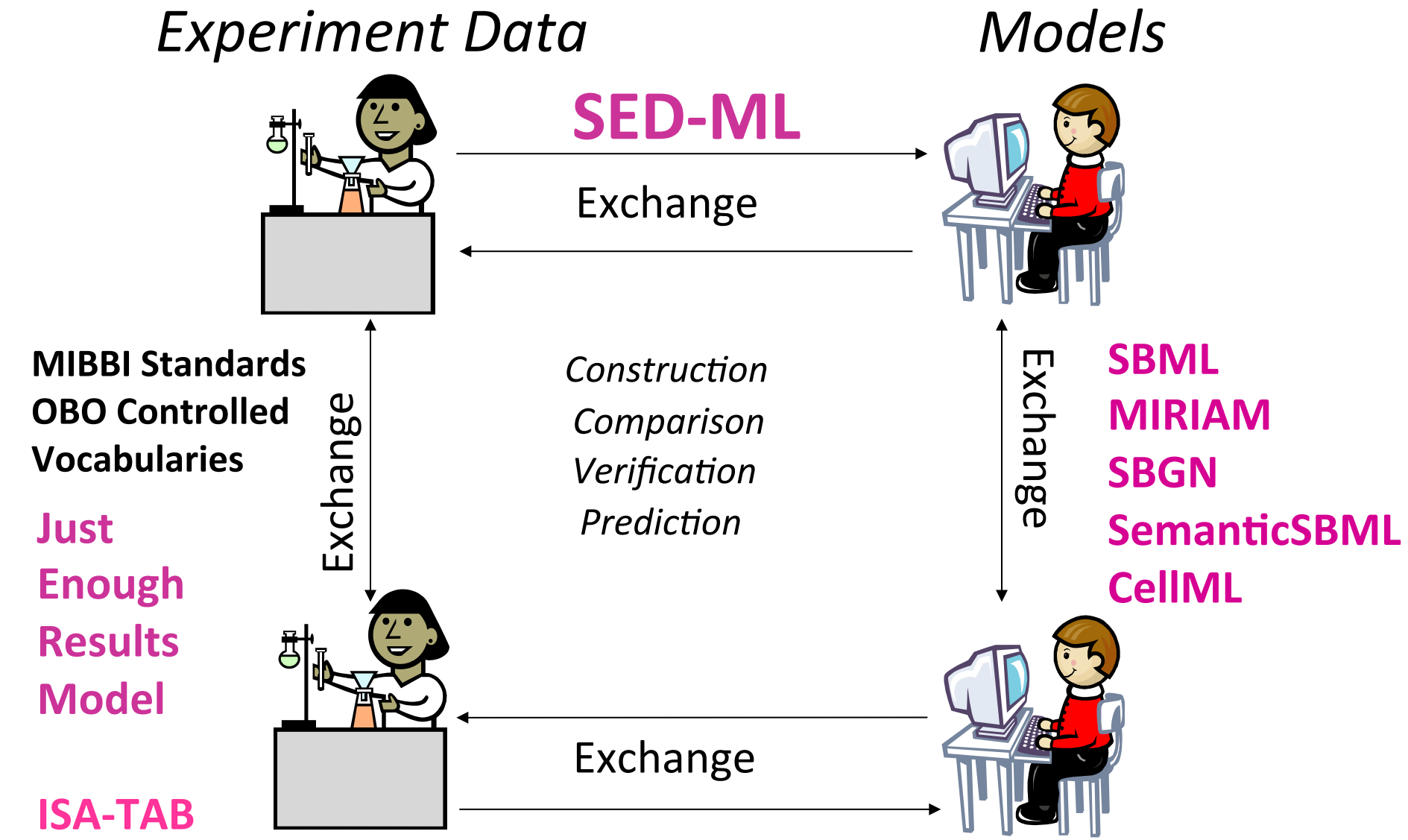








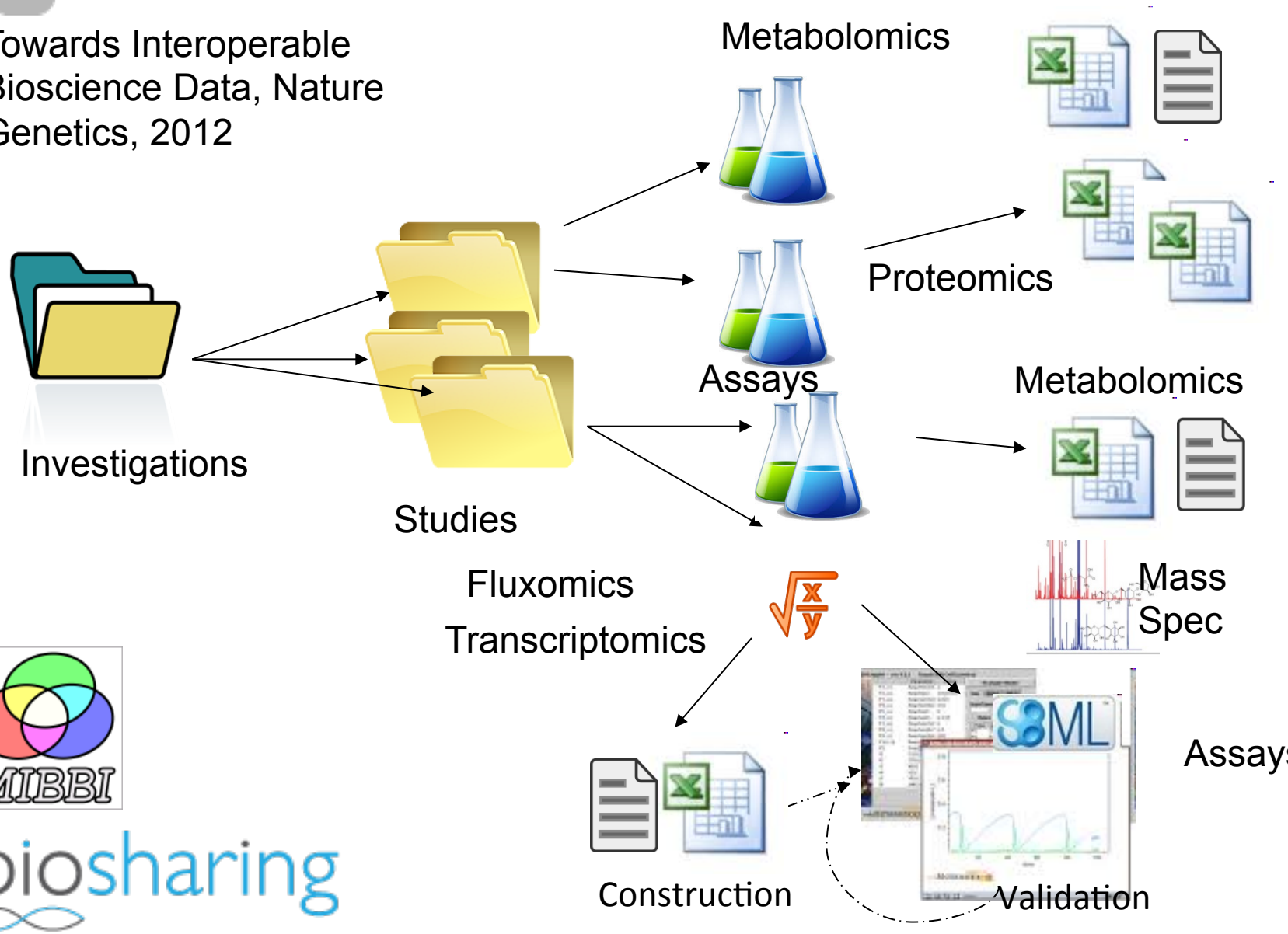
Standard Formats and Vocabularies



Simulation Experiment Description Markup Language

Standards, Structure, Interlink

Towards Interoperable Bioscience Data, Nature Genetics, 2012



MIBBI

biosharing

Find, share and exchange **Data**, **Models** and **Processes** within the SysMO Consortium.

Yellow pages | Experiment details | Assets | Activities | Documentation

Data files | Models | SOPs | Publications | Biosamples

All | External | Search

Log in / Register

Login:

Password:

Remember me:

Alternatively...

Log in using OpenID

Log in

Need an account?

Sign up

Forgotten password?

Tags [show all]

Bacillus subtilis Biochemi
Biochemistry and protein an
Bioinformatics Computati
and theoretical biology
Computational Systems Bio
Data Management dynamic
control of biological ne
Fermentation Genetic
modification Genetics
Mathematica Mathematic
modelling Matlab Metabolo
Microarray analysis
Microbiology
Molecular Biolog

Home > Data files Index

Latest A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Other All

Activation / Inhibition of 4 LDH enzymes in the presence of phosphate ions (Pi).

View
Download

Uploader: Anna Feldman-Salit
Creator: Anna Feldman-Salit
Other creators: *Not specified*

Version: 1
Projects: SysMO-LAB

The Table represents the simulation results of how the presence of phosphate ions (Pi) in the solution might affect the activity of four LDH enzymes. This includes the

Find, share and exchange **Data**, **Models** and **Processes** within the SysMO Consortium.

Yellow pages | Experiment details | Assets | Activities | Documentation

Data files | Models | SOPs | Publications | Biosamples

All

Log in / Register

Login:

Password:

Remember me:

Alternatively...

Log in using OpenID

Log in

Need an account?

Sign up

Forgotten password?

Tags [show all]

Bacillus subtilis Biochemistry
Biochemistry and protein analysis
Bioinformatics Computational
and theoretical biology
Computational Systems Biology
Data Management dynamics and
control of biological ne...
Fermentation Genetic
modification Genetics
Mathematica Mathematical

Home > Models Index > BSA115-sigB-adaptation

BSA115-sigB-adaptation

Download Model

The zip file contains model files and an experiment file. Unpack it in a directory and navigate with matlab to there. Use the 'matlab_execution_guide.m' for simulation and visualisation of the model. This file is written in matlab cell mode, so it is not a stand alone function.

Three models have been developed to test their capacity to reproduce the experimental data from Study: 'Controlled sigmaB induction in shake flask' with Assay: 'IPTG induction of sigmaB in BSA115'.
One model assumes a post-transcriptional inhibition (PTI) in the expression of the lacZ reporter gene. The PTI is either due to induced instability of lacZ mRNA or LacZ protein after activation of sigmaB. A hypothetical sigmaB dependent protein RegB is responsible for PTI. It is either a RNase or protease in this model.
One model assumes a transcription inhibitory effect on sigmaB dependent expression by a sigmaB dependent protein called RegB. This protein might bind to the sigma factor to decrease its efficacy in inducing transcription.
One model assumes that a protease called RegB acts specifically on sigmaB. This protease is also induced by sigmaB.
All models is based on mass-action (nearly so). Parameter for IPTG concentration is [100,200,1000], kbs, the expression parameter of Pspac-dependent sigmaB is fixed at 100. All other six parameters are estimated with two of the three experiments as training set (IPTG 0.1 and 0.2 mM) and one experiment as test set (IPTG 1mM). Parameter estimation is performed with SBToolbox2 using multiple rounds of particle swarm algorithm (pswarm).

1 item is associated with this Model:

Filename: BSA115-sigB-adaptation.zip

Uploa

sigB

Views: 19
Created:
Last update
Last used

Log in / Register

Login:

Password:

Remember me:

Alternatively...

Log in using OpenID

Log in

Need an account?

Sign up

Forgotten password?

Tags [show all]

Bacillus subtilis Biochemistry Biochemistry and protein analysis Bioinformatics Computational and theoretical biology Computational Systems Biology Data Management dynamics and control of biological ne... Fermentation Genetic modification Genetics Mathematica Mathematical modelling Matlab Metabolomics Microarray analysis Microbiology Molecular Biology Molecular biology techniques (RNA/DNA... Molecular microbiology ODE parameter estimation Proteomics Systems Biology Transcriptomics

Organisms

Bacillus subtilis
Clostridium acetobutylicum
Enterococcus faecalis
Escherichia coli
Lactic Acid Bacteria

Yeast Glycolytic Oscillations

Investigating oscillations at the level of yeast populations and individual cells

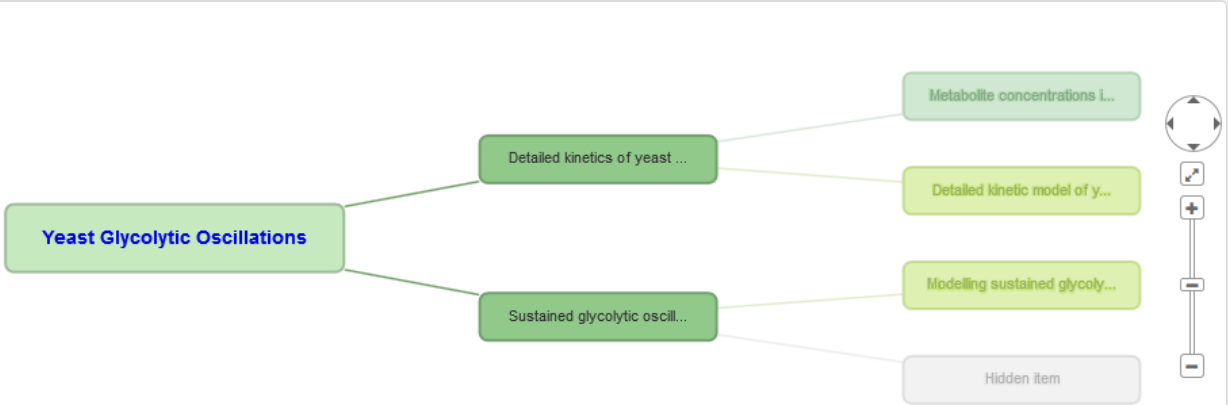
ID: 50

Projects: SysMO DB

Creators

Views: 136

Created: 20th Feb 2013 at 15:33



Chosen item

Investigation: Yeast Glycolytic Oscillations

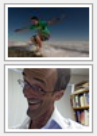
Connected items

Study: Sustained glycolytic oscillations in individual isolated yeast cells

Study: Detailed kinetics of yeast glycolytic oscillation

Related items

People (1) Projects (1) Studies (2) Assays (3+1) Data files (3+8) Models (10+1) Publications (3)



Das (1991) NADH and ATP concentrations determined during different phases of glycolytic oscillation in yeast extracts

View
Download

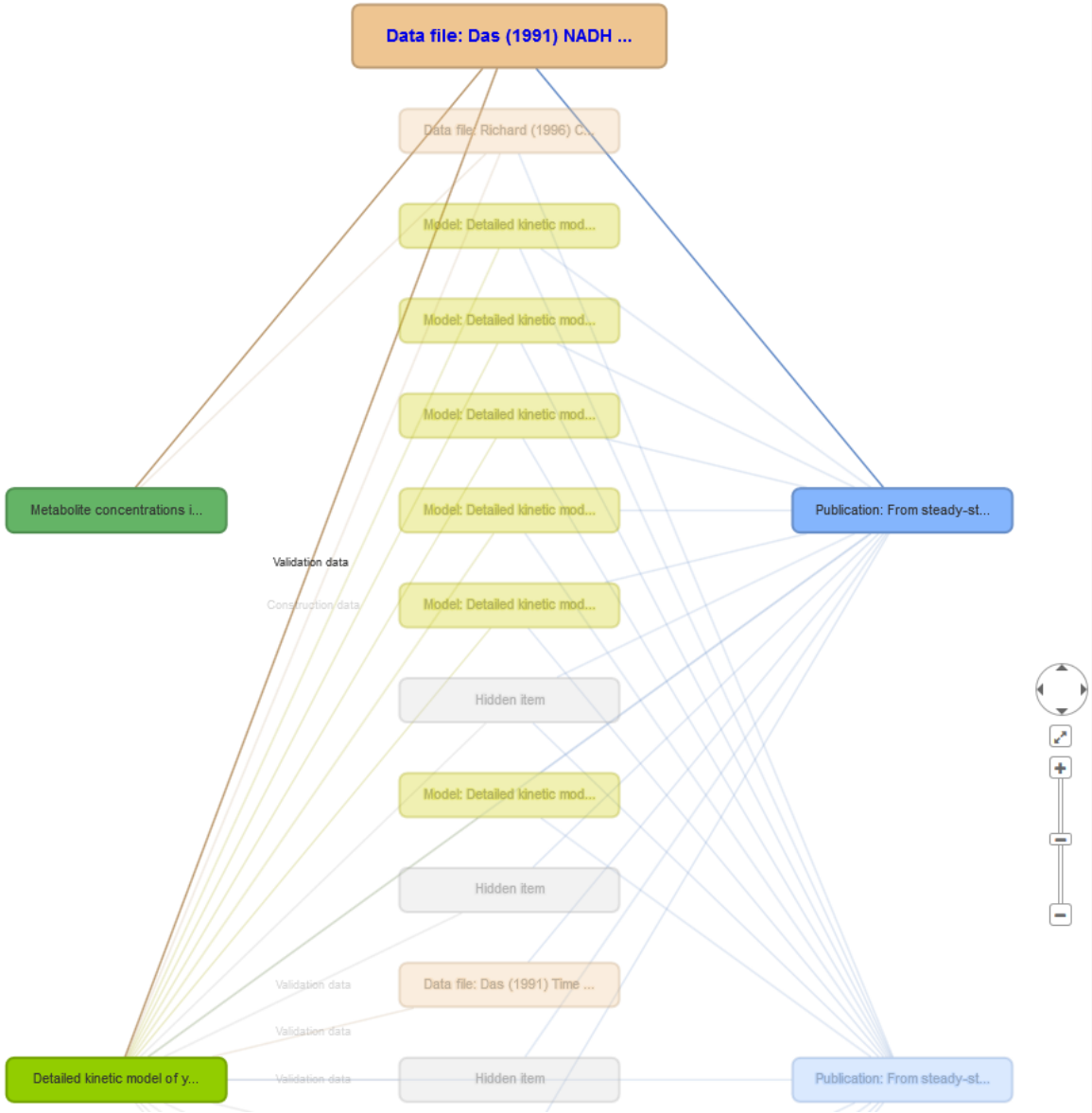
Uploader: Franco Du Preez
Creators: Franco Du Preez, Jacky Snoep

Version: 2
Projects: SysMO DB

modification Genetics Mathematica Mathematical modelling Matlab Metabolomics Microarray analysis **Microbiology** **Molecular Biology** Molecular biology techniques (RNA/DNA... Molecular microbiology ODE parameter estimation Proteomics Systems Biology Transcriptomics

Organisms

- Bacillus subtilis
- Clostridium acetobutylicum
- Enterococcus faecalis
- Escherichia coli
- Lactic Acid Bacteria
- Lactobacillus plantarum
- Lactococcus lactis
- Pseudomonas fluorescens
- Pseudomonas putida
- Saccharomyces cerevisiae
- Streptococcus pneumoniae
- Streptococcus pyogenes
- Streptomyces coelicolor
- Sulfolobus solfataricus
- Trypanosoma brucei



Chosen item

Data file: Das (1991) NADH and ATP concentrations determined during different phases of glycolytic oscillation in yeast extracts

Connected items

Publication: From steady-state to synchronized yeast glycolytic oscillations II: model validation
Modelling Analysis: Detailed kinetic model of yeast glycolytic oscillation
Experimental Assay: Metabolite concentrations in yeast glycolytic oscillations

- Asset
- Equipment
  - Instrument
  - Software
- Experiment_description
  - Assay
  - Investigation
  - Study
- Experimental_factors
  - Assay_type
  - Bio_material
    - Sample
    - Specimen
  - Characteristic_for_specimen_or_sample
  - Factors_studied
  - Measured_item
    - Biological_entity
  - Microarray_factors
  - Proteomics_factors
    - Ionization_source
    - Normalization_method
    - Post_source_component
    - Quantitation
    - Spectrum_generation
    - Spectrum_interpretation
  - Technology_type
- Format
  - Data_file_format
  - Model_execution_environment
  - Model_format
  - SOP_file_format
- Person
- Project

262 classes and 43 properties

Just Enough Results Model

- Overarching Application Ontology
- Describes and enriches the relationships between things produced and used in experiments
  - experiment descriptions
  - provenance of projects
  - data, models, results, samples, protocols, standard operating procedures, publication…
- For the experiments themselves
  - reuse community ontologies, markups, mim, identifiers
  - MGED/EFO, PSI-MS, SBO, CheBI ids etc

http://bioportal.bioontology.org/ontologies/JERM expressed in OWL

http://bioportal.bioontology.org/ontologies/JERM

Type of data	What was measured	Values in the datasets	Biological sample and treatments applied
Microarray, growth curve, enzyme activity…	Gene expression, OD, metabolite concentration…	Units, time series, repeats…	

Common elements

reuse community ontologies, markups, mim, identifiers

Enzymic reactions
reactions catalyzed, substrates, products, inhibitions
CheBI ids

Microarray
QC methods
normalisation
MGED/EFO

Proteomics
instruments
PSI-MS

Data type specific elements

Just **E**nough **R**esults **M**odel
Describes and enriches the relationships between things produced and used in experiments.

http://bioportal.bioontology.org/ontologies/JERM

isa

Type of data	What was measured	Values in the datasets	Biological sample and treatments applied
Microarray, growth curve, enzyme activity…	Gene expression, OD, metabolite concentration…	Units, time series, repeats…	
		Common elements	

JERM

Just **E**nough **R**esults **M**odel
Describes and enriches the relationships between things produced and used in experiments.

reuse community ontologies, markups, mim, identifiers

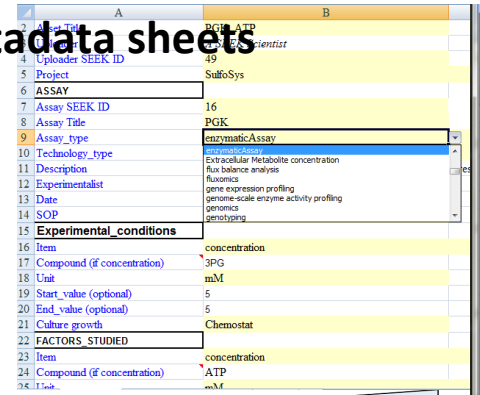
Enzymic reactions
reactions catalyzed, substrates, products, inhibitions
CheBI ids

Microarray
QC methods
normalisation
MGED/EFO

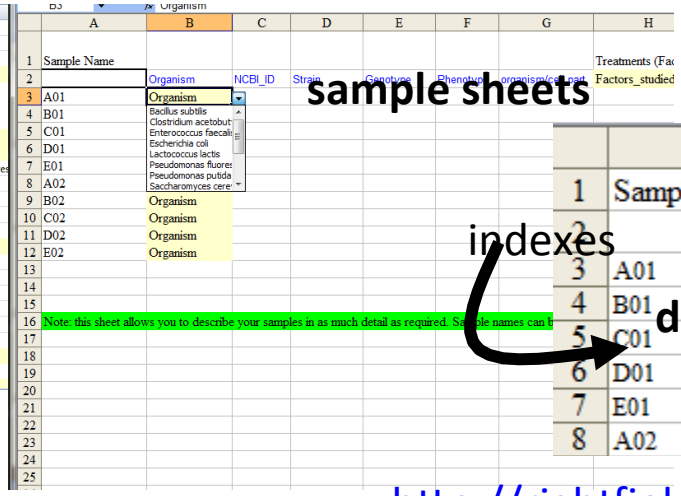
Proteomics
instruments
PSI-MS

Data type specific elements

metadata sheets

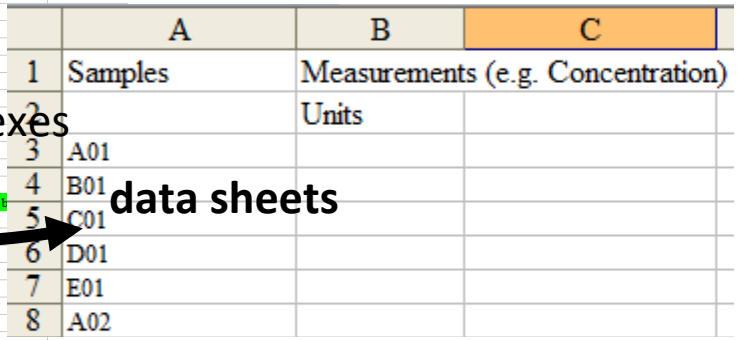


sample sheets



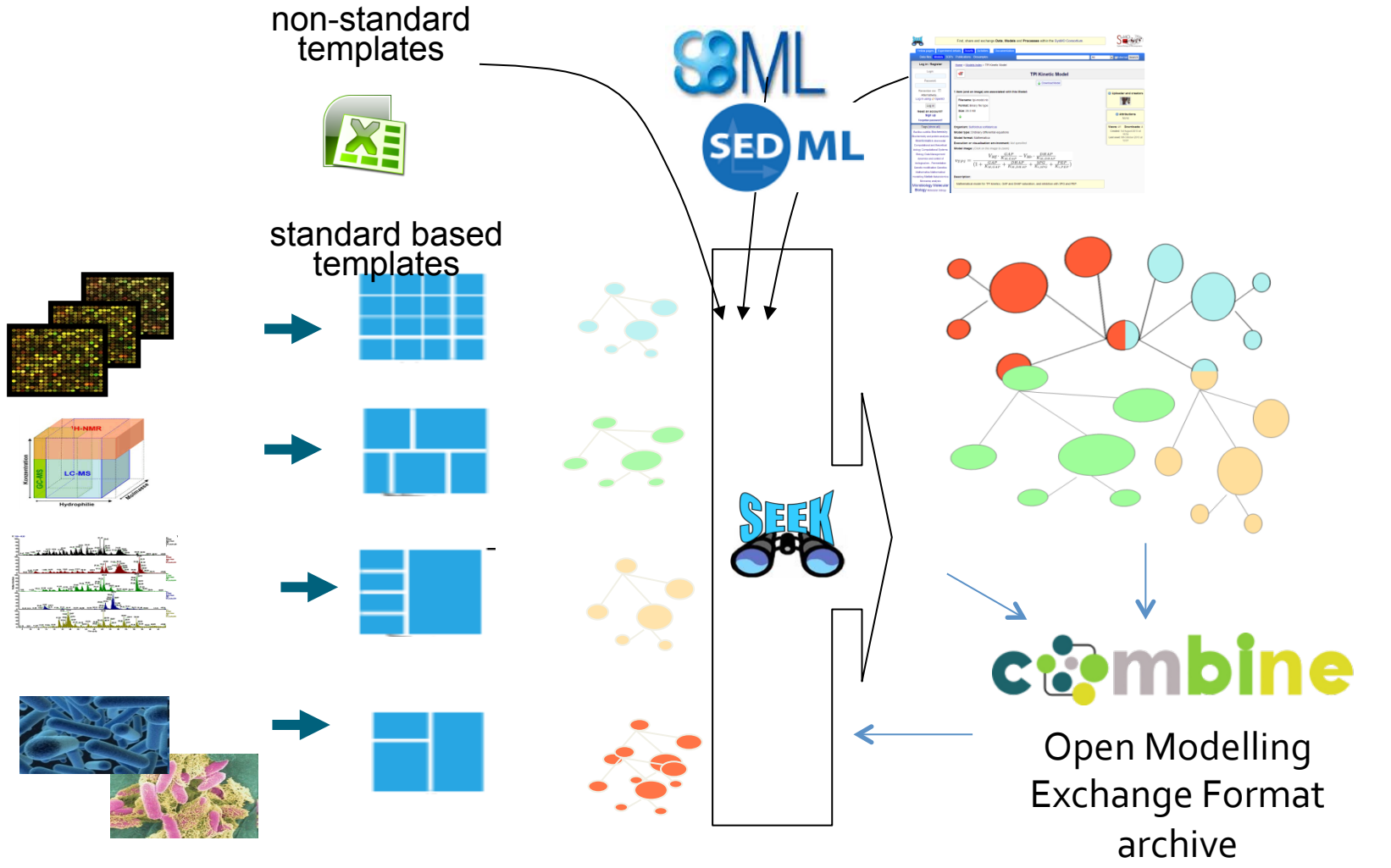
indexes

data sheets



RightField

http://rightfield.org.uk/



Different types of data
Plugins to registered data repositories

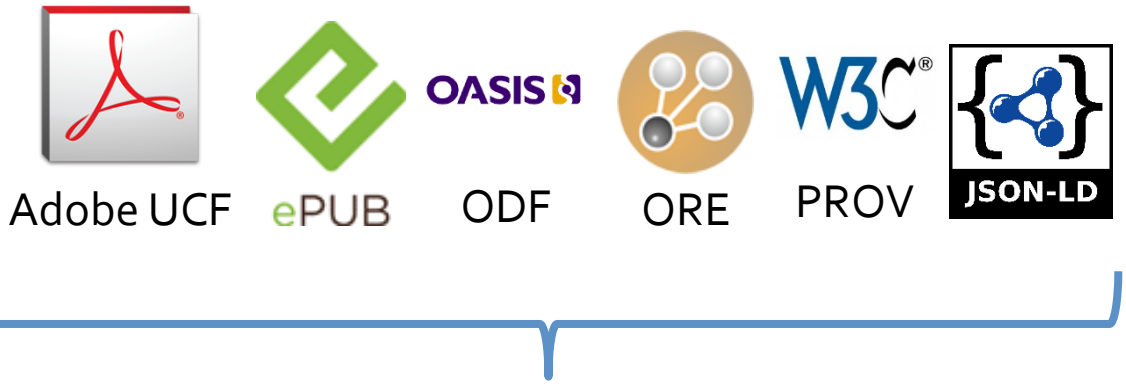
Extract and auto-catalogue metadata

Define relationships, cross-link, aggregate, query

Sys Bio Research Objects

http://www.researchobject.org

Systems Biology:
A common archive format for reuse across tools



combine

OMEX archive

researchobject

Research Object Bundle

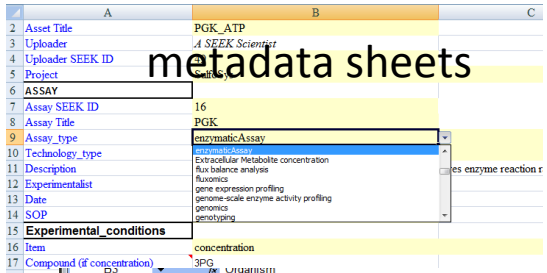
- Aggregation
- Annotations/provenance
- Ad-hoc domain-specific specification

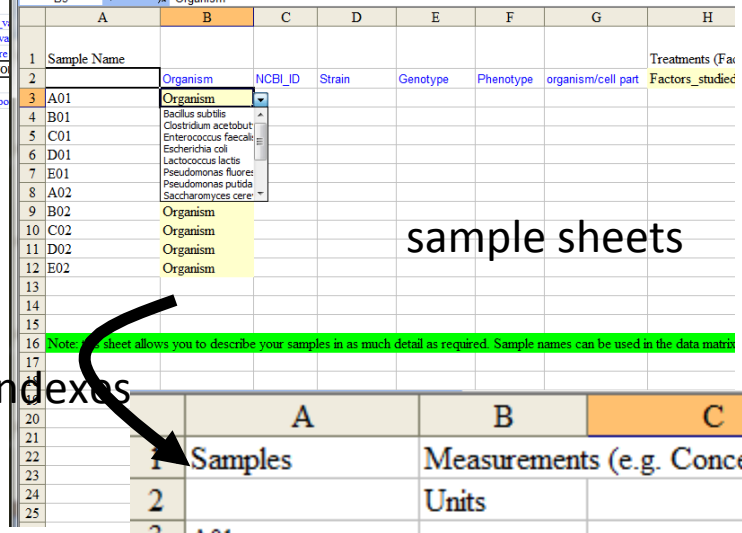
Workflow 4Ever

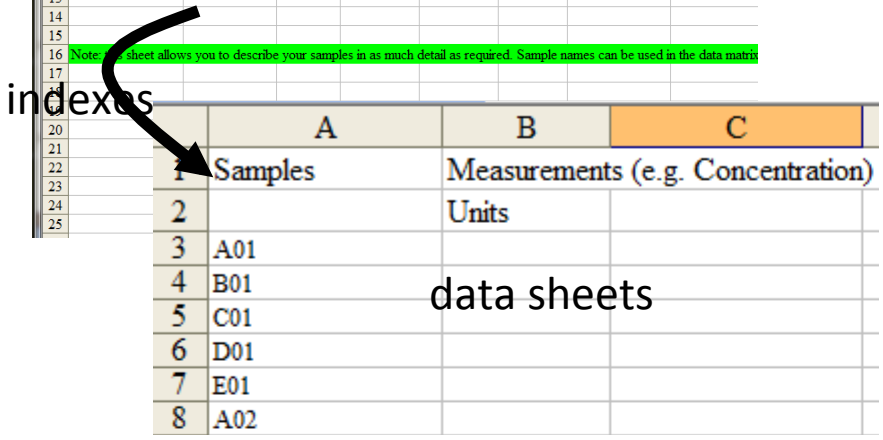
RightField

http://rightfield.org.uk/

Semantic Annotation by Stealth Instrumented Spreadsheets





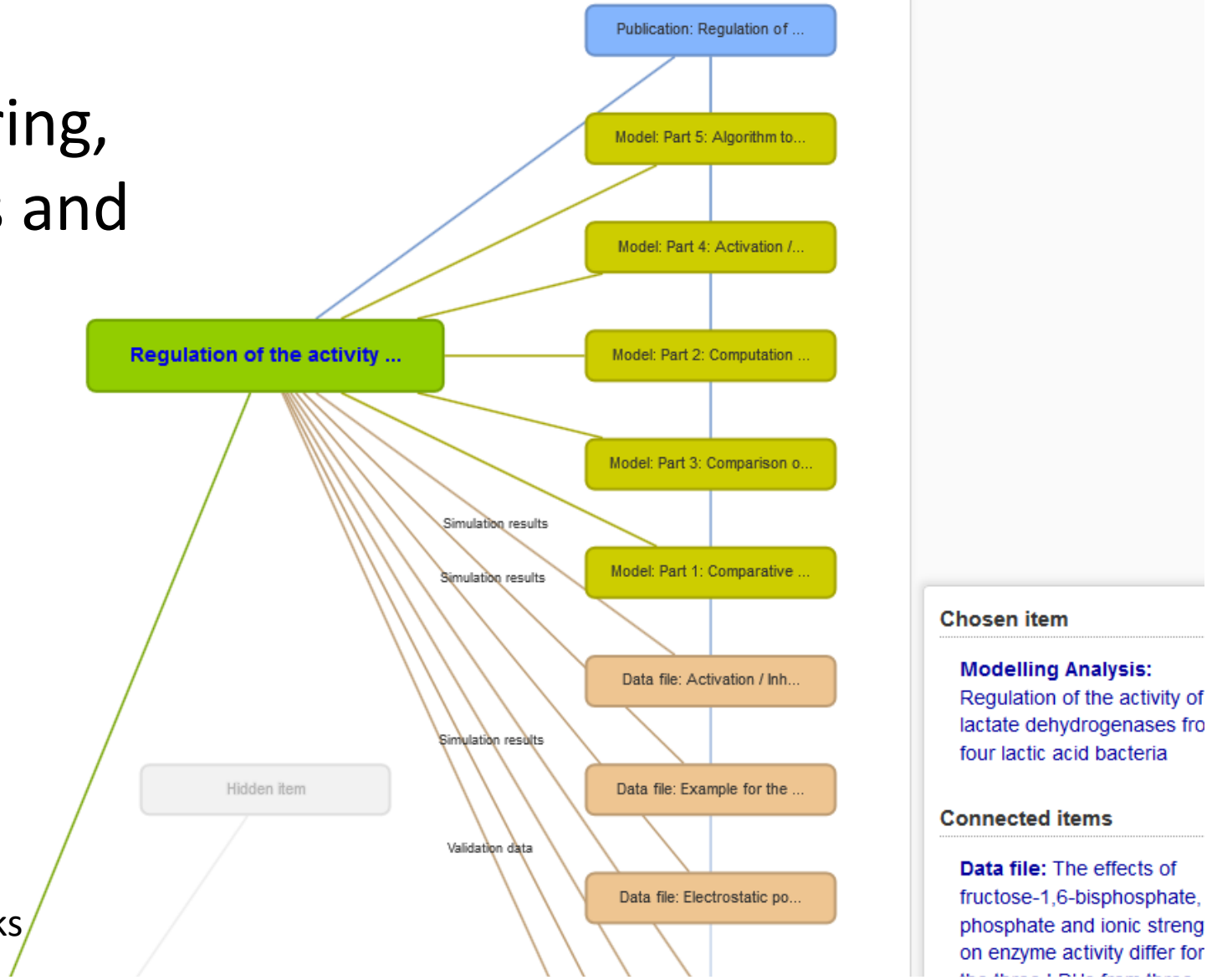


- Desktop tool
- Embed ranges of ontology terms into spreadsheets
- Annotate data with terms through drop-down lists
- Automatically extract semantically annotated metadata from annotated spreadsheets (and produce RDF)

Question	Lucene	RDF
Which experiments were carried out on *E.coli* (organism X)?	+	+
What proteomic (experiment x) data is available? What types of transcriptomics (assay type x) experiments were performed?	+/+	+/+
Who has experimental data on gene/protein/metabolite X	+/+/+	+/+/+
What data was used to construct the model and what data was used to validate it?	+	+
Who is in the COSMIC (project X) project?	+	+
Are there any models on yeast (models on X)?	+	+
What SOP were used in experiment X? are there any protocols for Mass Spectroscopy (tech type X) experiments?	-	+
Which microarray data files show up-regulation in genes with Gene ontology molecular function X	-	+
Who is in more than one SysMO project?	-	+
What are the factors studied in the MOSES project (project X)?	-	+
What range of concentrations of metabolites (extra- and intracellular) are detectable from organism XXX	-	+
How good is the correlation between transcriptome levels, proteome levels and enzyme activities in organism X in study Y? Is a time delay observed?	-	-

Reproducible (Open?) Research

Data sharing, openness and careers incentive



See Titus and Phil talks

Open Research: Research Groups & Lifecycles

- Sharing policy
- Visibility, Downloadability
- Fine grained permissions

- Protocols for
  - Management transfer
  - Visibility feedback and sharing workflows
  - Publication data deposition in external public stores
  - Batch publishing

Owner

Within Project

Versions

Retractions

Across Projects

Versions

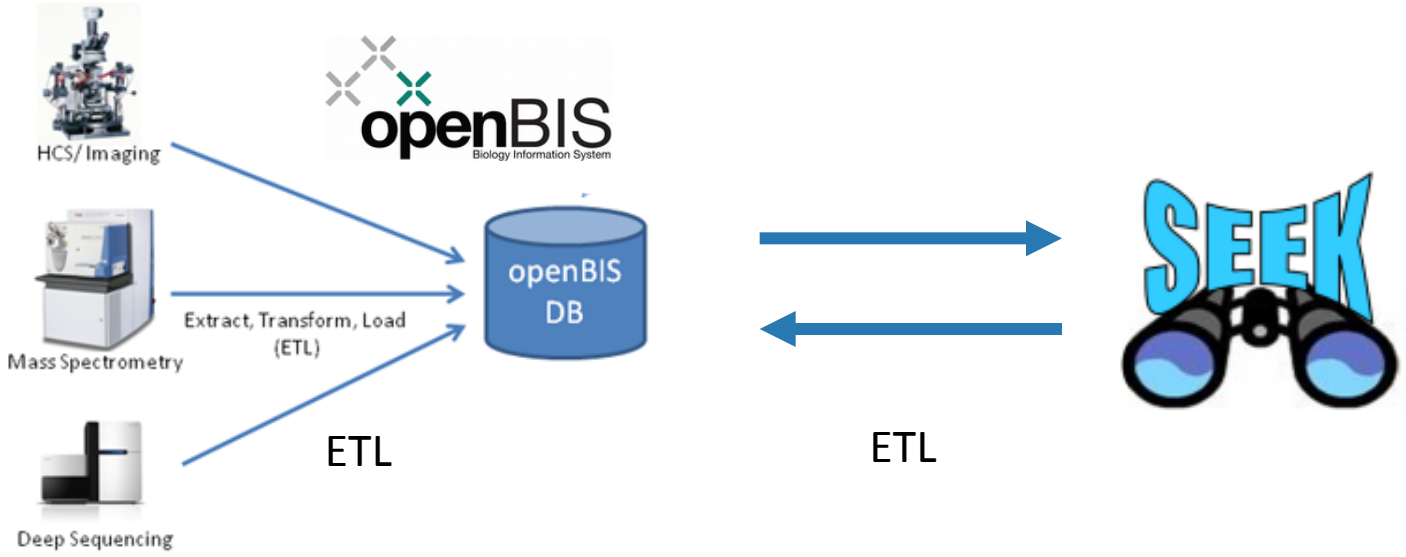
Public

Final version
No Retraction

Manager

Gatekeeper

OpenBIS and SEEK



Open BIS back-end: Extract, Transform and Load tooling direct from the instrumentation, along with data analysis pipelines. Automatic archiving. Handles large data.

SEEK front-end: A rich web interface for describing, finding, linking and promoting ongoing research and outcomes.

Open Source Customisable Platform

https://bitbucket.org/seek4science/seek

Yeast Glycolysis

unicellsys

WEL mcr Web Ergonomics Lab

Vrije Universiteit, Amsterdam

GLYCON

sybacol

EVIMalaR

CISBIO

ElasticAP

RosAge
Reactive oxygen species and the dynamics of ageing

eagle genomics

Systems Science for Health (SSfH)

FORSYS MACS

SBCancer.

SulfoSYS biotec

Briefings in Bioinformatics Advance Access published October 9, 2012
BRIEFINGS IN BIOINFORMATICS. page 1 of 14 doi:10.1093/bib/bbs064

Data management strategies for multinational large-scale systems biology projects

Wasco Wruck, Martin Peuker and Christian R.A. Regenbrecht

Submitted: 13th July 2012; Received (in revised form): 4th September 2012

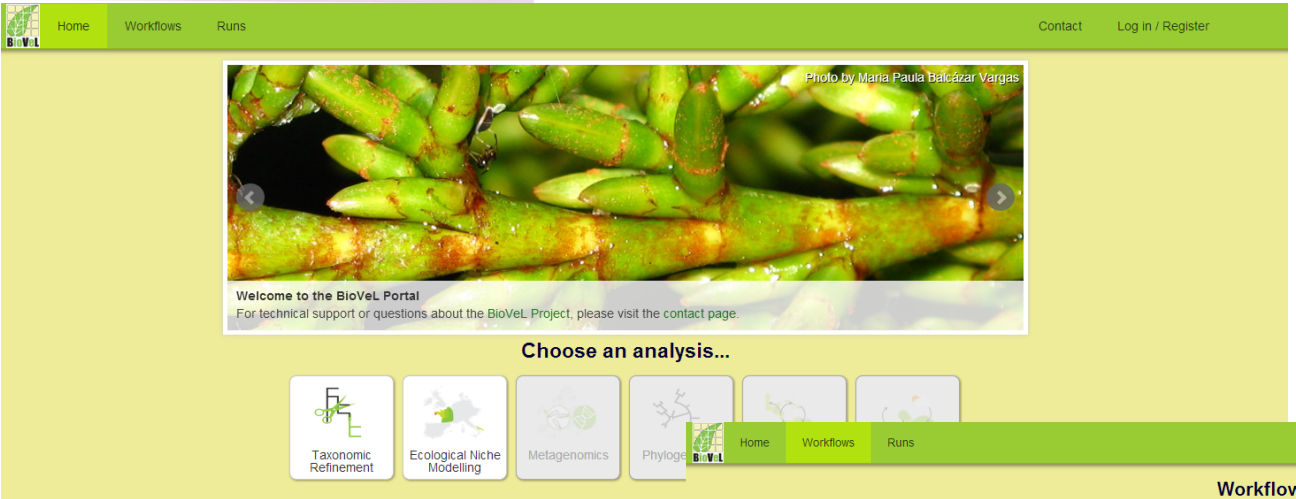
Key Points

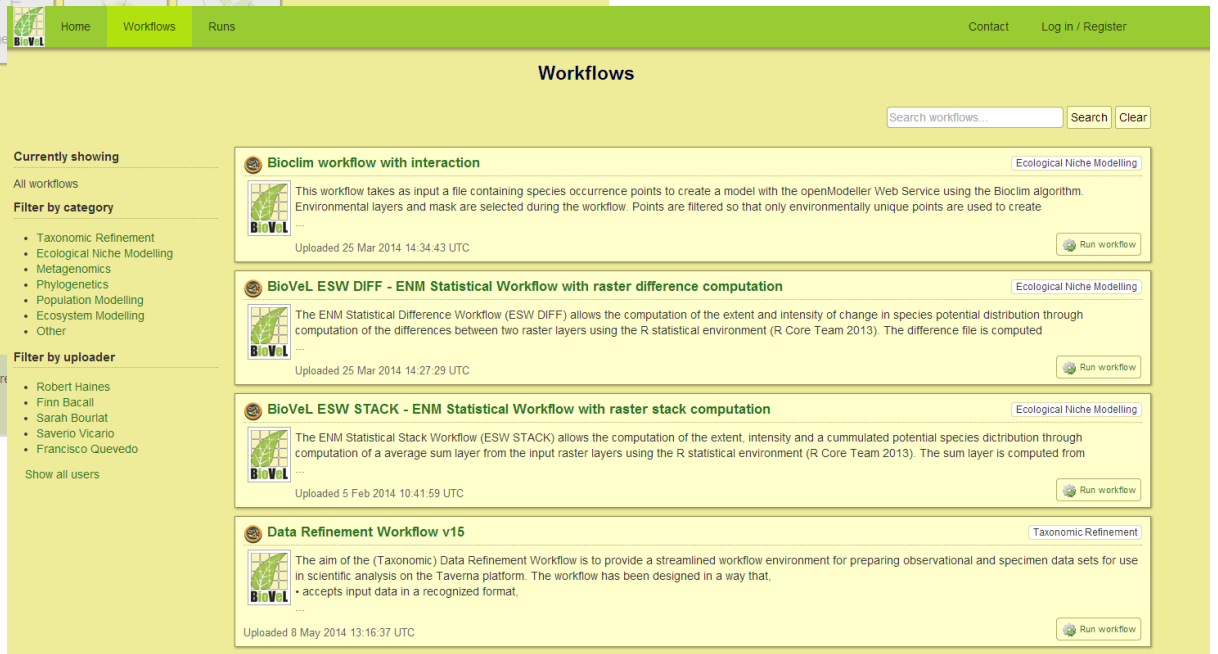
- Incentives to share data can be given by data citation credits (datacite).
- Open access to research data can be advanced via making sharing a condition of funding.
- Data management systems might be made attractive via alleviating and improving researchers' work, e.g. support for standard formats and publications.
- The reviewed systems proved useful for systems biology projects at least in dedicated environments, SysMO-SEEK out-of-the-box provides most useful features for large-scale systems biology projects.

http://bib.oxfordjournals.org/content/early/2012/10/09/bib.bbs064.full.p

Open Source Customisable Platform

https://bitbucket.org/seek4science/seek





http://fair-dom.org/

Open Facility for European Systems Biology data & model management

seeded by EU programmes

- **Platform**
  - SEEK + openBIS + new features & styling
- **Resource**
  - EuroSEEK + pool of community resources (including established SEEKs).
  - Independent researchers. Secure data.
- **Facility**
  - Curation & support services, training

BBSRC
bioscience for the future

Federal Ministry of Education and Research

SystemsX.ch
The Swiss Initative in System Biology

NWO

SEEK
openBIS
Biology Information System

http://fair-dom.org/

Open Facility for European Systems Biology data & model management

seeded by EU programmes

- **Community**
  - workshops, user and developer forums, knowledge network, standards & policy, training, FAIRDOM Foundation, **Model Carpentry.**
- Sys Bio Developers Foundry workshop

  6-7 October Heidelberg

  http://fair-dom.org/wiki/Foundry_workshop
- **RI**
  - working with other EU RIs, an EU network of national facilities, funding models.

BBSRC
bioscience for the future

Federal Ministry of Education and Research

SystemsX.ch
The Swiss Initative in System Biology

NWO

SEEK
openBIS
Biology Information System

MANCHESTER 1824
The University of Manchester

Carole Goble

Natalie Stanford

Stuart Owen

also contributing:
UK SEEK team

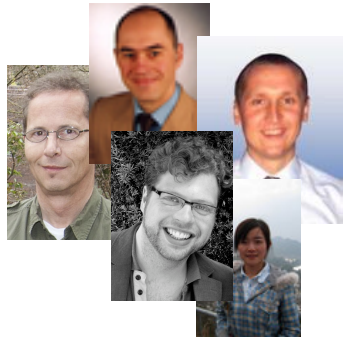
Jacky Snoep

HITS

Wolfgang Mueller

Olga Krebs

Quyen Nguyen



also contributing:
VLN SEEK team

SystemsX.ch
The Swiss Initiative in Systems Biology

Peter Kunszt

Bernd Rinn

Katy Wolstencroft