Heidelberg Institute for Theoretical Studies

HITS

Systems biology requirements for standardisation and integration of wet and dry laboratory data.

Olga Krebs, Heidelberg Institute for Theoretical Studies
Heidelberg, Germany

THE YOUNG SCIENTISTS SCHOOL, Novosibirsk, 28-29 June 2010

Outline

- Overview of systems biology
- Introduce some important concepts
- Standards, controlled vocabularies and metadata
    - wet lab
    - dry lab
- Tools, databases

Systems Biology Has its Backers and Attackers

Though coined 40 years ago, a lot of people still ask, "What's that?" when the term systems biology comes up.

"It is used in so many different contexts, nobody is really clear what you mean by it," says John Yates , a professor at the Scripps Research Institute.

David Placek, president of Sausalito: "Systems biology is just so general that it could apply to many things. When you're naming a category, the underlying principle is that if you make a statement like, 'I'm doing systems biology,' do people know what you're talking about?"……

Volume 17 | Issue 19 | 27
Oct. 6, 2003, The Scientist

What Is Systems Biology?

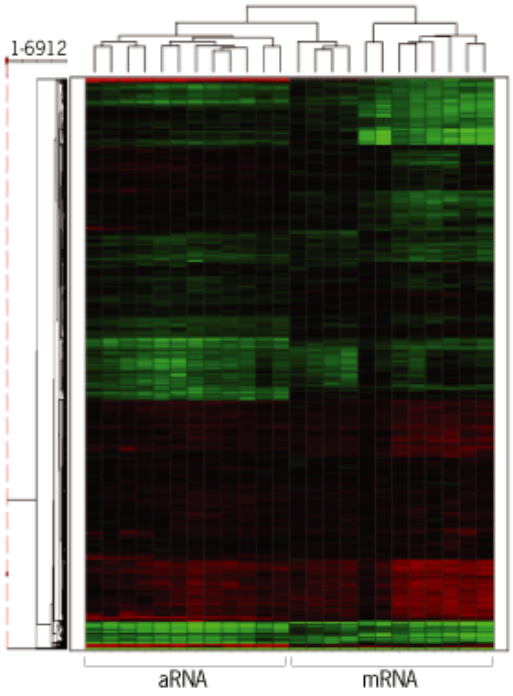
Biology went top-down for the last 50 years

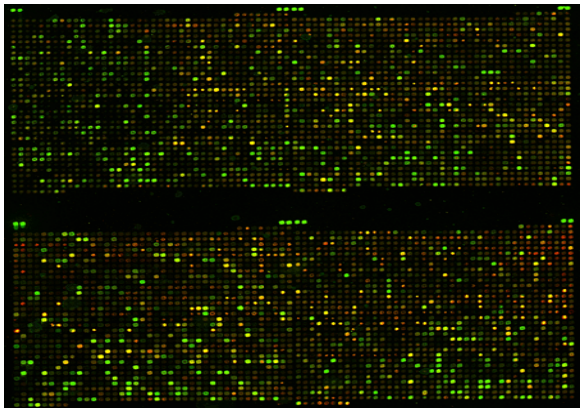
- From cell to protein to gene ..
- Huge amounts of data produced

Challenge: put the pieces back together again

Systems Biology?

High-throughput Data?





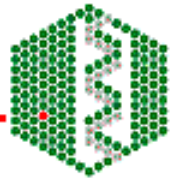
Current systems biology craze has resulted from recent technologies that allow for rapid or simultaneous measurement of large amounts of biomolecular data (e.g. genomics, microarrays, etc.)

Systems Biology?

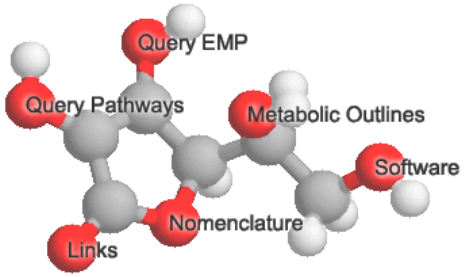
Databases?

SigPath institute for computational biomedicine icb

EMBL Outstation
European Bioinformatics Institute

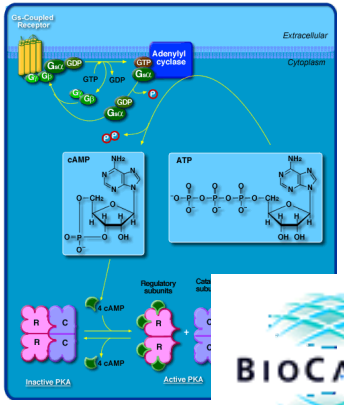


BRENDA



NCGR

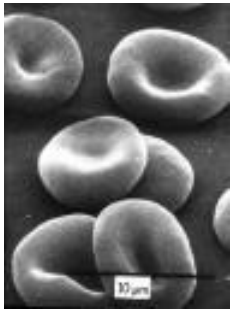
PathDB



BIOCARTA

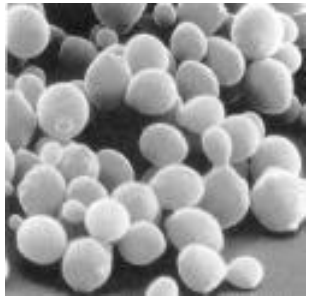
.. or is it modelling?

Red Blood Cell



Mulquiney, Joshi, Heinrich, …

Yeast Glycolysis

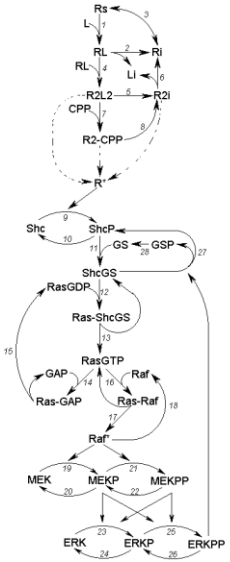


Bas Teusink

Trypanosoma Brucei

Barbara Bakker, Westerhoff and Cornish-Bowden

EGF Signaling Pathway

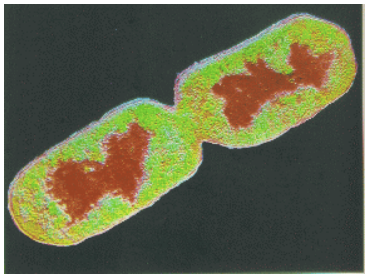


Frances Brightman et al

Calvin Cycle

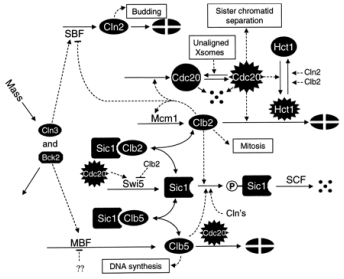
Poolman and Fell

Chemotaxis, ecoli



Many Contributors

Yeast Cell Cycle



John Tyson et al

Systems Biology - *The study of the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology involves (1) collection of large sets of experimental data (2) proposal of mathematical models that might account for at least some significant aspects of this data set, (3) accurate computer solution of the mathematical equations to obtain numerical predictions, and (4) assessment of the quality of the model by comparing numerical simulations with the experimental data.*

First described in 1999 by Leroy Hood

What Is Systems Biology?

Systems biology is devoted to a new science, a critical science of the future that seeks to understand the **integration** of the pieces to form biological systems.

David Baltimore, Nobel Laureate

Systems Biology- bridging the culture gap?

Molecular biologists are deluged with data, and physicists, used to reducing complex systems to basic principles, might help to make sense of it all.

But bringing the two disciplines together isn't easy, says Jonathan Knight.

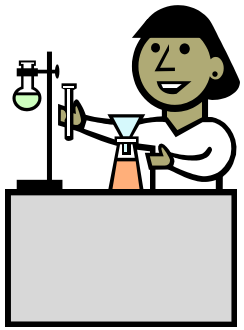
Dissatisfaction With Biologists

- Frustration at the failure of biologists to engage with quantitation, formal models, theory and to predict how overall behaviour arises out of the interaction of the components.
- A belief that only physicists, engineers, mathematicians, computer scientists can take on these issues.

Systems biology needs communication

Experimentalists

Modellers

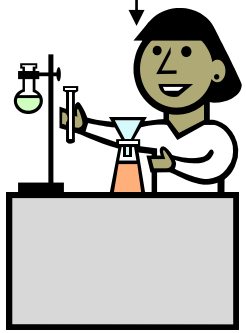


Exchange

Exchange

Bioinformaticians

Exchange



Exchange

RESEARCH ARTICLE

Global Mapping of the Yeast Genetic Interaction Network

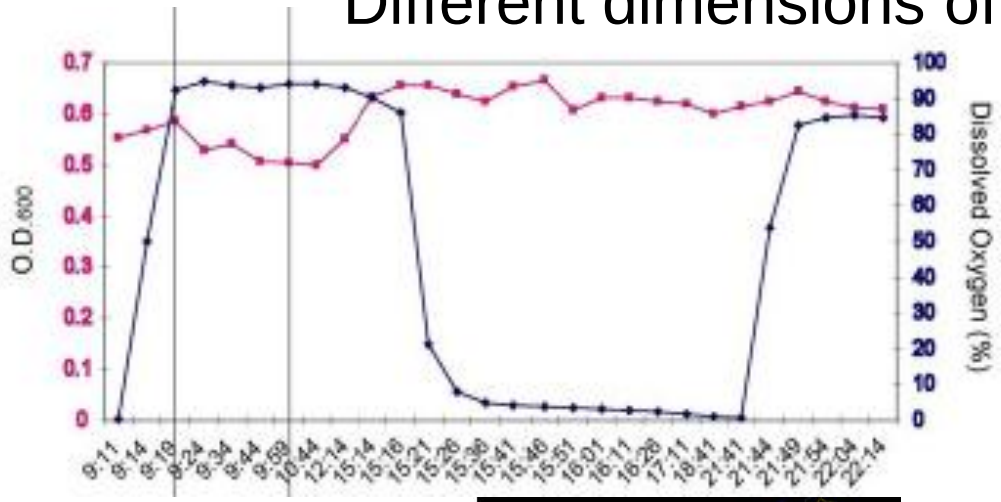
Amy Hin Yan Tong,[1,2]* Guillaume Lesage,[3]* Gary D. Bader,[4] Huiming Ding,[1] Hong Xu,[1,2] Xiaofeng Xin,[1,2] James Young,[6] Gabriel F. Berriz,[7] Renee L. Brost,[1] Michael Chang,[5] YiQun Chen,[1] Xin Cheng,[1] Gordon Chua,[1] Helena Friesen,[2] Debra S. Goldberg,[7] Jennifer Haynes,[2] Christine Humphries,[2] Grace He,[1] Shamiza Hussein,[3] Lizhu Ke,[1] Nevan Krogan,[1,2] Zhijian Li,[1,2] Joshua N. Levinson,[3] Hong Lu,[1] Patrice Ménard,[3] Christella Munyana,[3] Ainslie B. Parsons,[1,2] Owen Ryan,[1] Raffi Tonikian,[1,2] Tania Roberts,[5] Anne-Marie Sdicu,[3] Jesse Shapiro,[3] Bilal Sheikh,[1] Bernhard Suter,[8] Sharyl L. Wong,[7] Lan V. Zhang,[7] Hongwei Zhu,[1] Christopher G. Burd,[9] Sean Munro,[10] Chris Sander,[4] Jasper Rine,[8] Jack Greenblatt,[1,2] Matthias Peter,[11] Anthony Bretscher,[6] Graham Bell,[3] Frederick P. Roth,[7] Grant W. Brown,[5] Brenda Andrews,[2]† Howard Bussey,[3]† Charles Boone[1,2]†

52 authors from 11 institutions

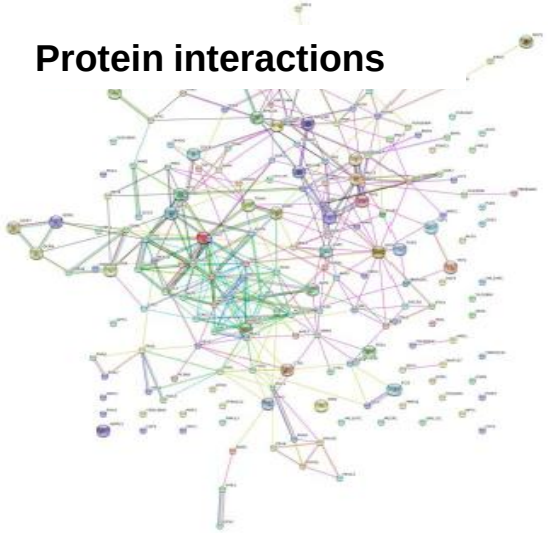
A genetic interaction network containing ~1000 genes and ~4000 interactions was mapped by crossing mutations in 132 different query genes into a set of ~4700 viable gene yeast deletion mutants and scoring the double mutant progeny for fitness defects. Network connectivity was predictive of function because interactions often occurred among functionally related genes, and similar patterns of interactions tended to identify components of the same pathway. The genetic network exhibited dense local neighborhoods; therefore, the position of a gene on a partially mapped network is predictive of other genetic interactions. Because digenic interactions are common in yeast, similar networks may underlie the complex genetics associated with inherited phenotypes in other organisms.

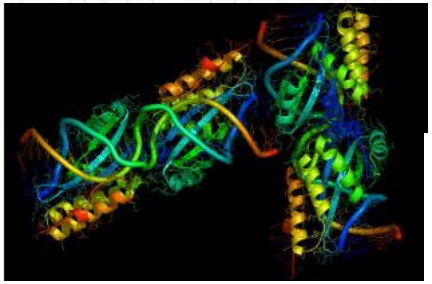
What are we dealing with?

Different dimensions of systems biology data

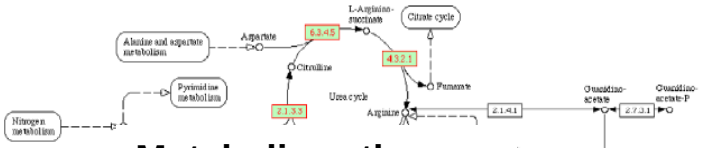


Protein interactions

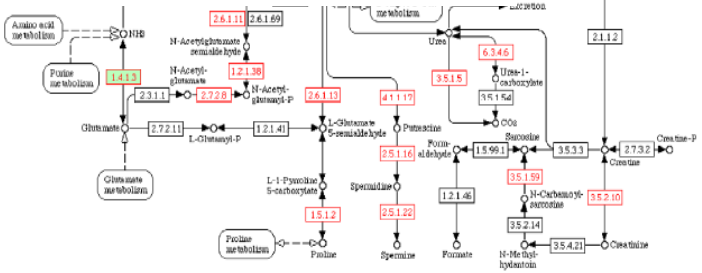




Functional anotations



Metabolic pathways



Microarray data

nature

www.nature.com/nature

Vol 461 | Issue no. 7261 | 10 September 2009

Data's shameful neglect

Research cannot flourish if data are not preserved and made accessible.

More and more often these days, a research project's success is measured not just by the publications it produces, but also by the data it makes available to the wider community.

Nature has a new data sharing policy . .

- "Before submitting the paper, at least one senior member from each collaborating group must take responsibility for their group's contribution.
- “Three major responsibilities are covered:
  - preservation of the original data on which the paper is based,
  - verification that the figures and conclusions accurately reflect the data collected and that manipulations to images are in accordance with Nature journal guidelines, and
  - minimization of obstacles to sharing materials, data and algorithms through appropriate planning.“

Why Standards

Standardization is not removing diversity but improving connection, documentation, annotation and scalability, it is a crucial step forward to data interoperability – data have to be stored, exchanged and re-used

...to enable an unambiguous exchange, comparison, integration and interpretation of the data

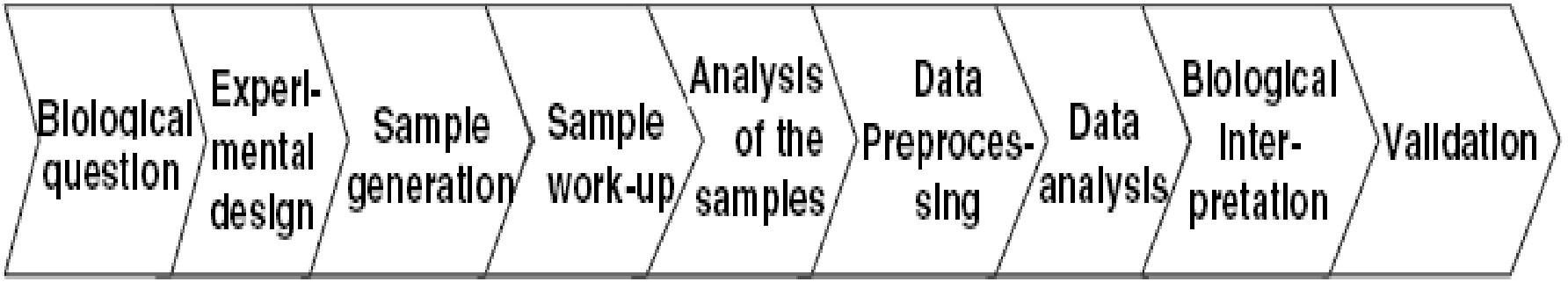
- Communication needs shared **standards** (language, grammer, vocabulary, ...)

- To avoid misunderstanding **definitions** (What is a compound? gene? protein?)

A **controlled vocabulary** is unambiguously defined and standardised

→ You have to speak the same language to communicate

Experimental workflow



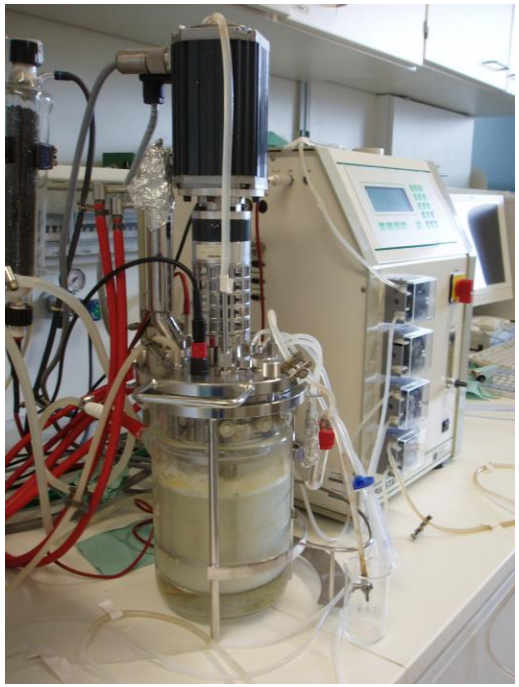
Standardised Experimental Setup

C. acetobutylicum strain ATCC 824

- Continuous culture in a chemostat
- Phosphate limitation
- Cells are harvested at steady state

Constant parameters:

- 4% glucose in medium
- Dilution rate
- Temperature 37 °C



It is adjusted to constant values:

- Acids: 5.7 pH
- Solvents: 4.5 pH

What is measured:

- Metabolites & enzymes
- Protein concentrations
- Expression rates

Defined standard operating procedures (SOP) for extracting and handling of different types of samples!

SOP – Standard Operating Procedure

- Sharing experimental protocols
- Agreement within projects on strains and conditions
- Good scientific practice
- Understanding data, experiments and results for verification or for modifying for use in other experiments

Standard Operation Procedures

SOP - Public Initiatives

- Open Wetware – public access, no recommended format
- Nature Protocols – published SOPs with links to publications – defined format
- Cold Spring Harbour Protocols – proprietary published SOPs – defined format
- SySMO DB

Metabolomics (2007) 3:189–194
DOI 10.1007/s11306-007-0080-4

ORIGINAL ARTICLE

Standard reporting requirements for biological samples in metabolomics experiments: microbial and in vitro biology experiments

Mariët J. van der Werf · Ralf Takors · Jørn Smedsgaard · Jens Nielsen · Tom Ferenci · Jean Charles Portais · Christoph Wittmann · Mark Hooks · Alberta Tomassini · Marco Oldiges · Jennifer Fostel · Uwe Sauer

Received: 3 January 2007 / Accepted: 23 July 2007 / Published online: 20 August 2007

Abstract With the increasing use of metabolomics as a means to study a large number of different biological research questions, there is a need for a minimal set of reporting standards that allow the scientific community to evaluate, understand, repeat, compare and re-investigate metabolomics studies. Here we propose, a first draft of minimal requirements to effectively describe the biological context of metabolomics studies that involve microbial or in vitro biological subjects. This recommendation has been produced by the microbiology and in vitro biology working subgroup of the Metabolomics Standards Initiative in collaboration with the yeast systems biology network as part of a wider standardization initiative led by the Metabolomics Society. Microbial and in vitro biology metabolomics is defined by this sub-working group as studies with any cell or organism that require a defined external medium to facilitate growth and propagation. Both a minimal set and a best practice set of reporting standards for metabolomics experiments have been defined. The minimal set of reporting standards for microbial or in vitro biology metabolomics experiments includes those factors that are *specific* for metabolomics experiments and that critically determine the outcome of the experiments. The best practice set of reporting standards contains both the factors that

The Microarray Gene Expression Data Society (MGED)

MGED is a group of researchers with the intention of establishing standards for microarray data annotation and to enable the creation of public databases for microarray data.

MGED's work is arranged into four working groups:

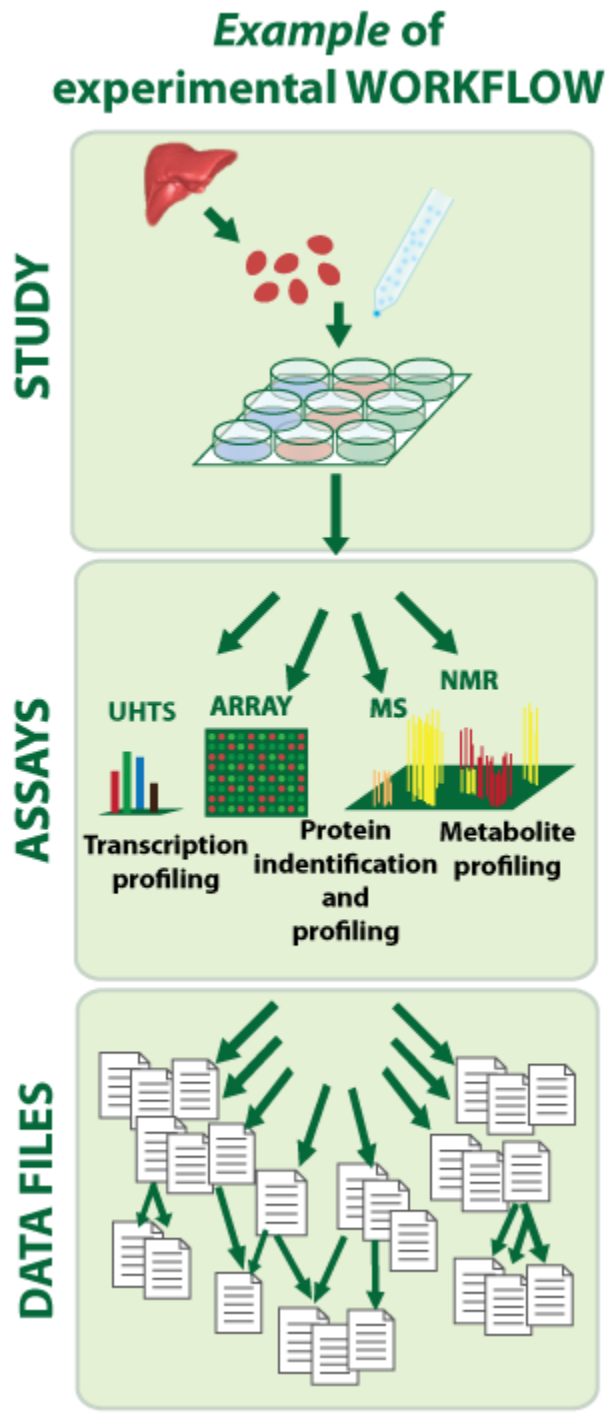
- *MIAME*. Minimal Information About a Microarray Experiment. Formulates the information required to record about a microarray experiment in order to be able to describe and share the experiment.
- *Ontologies*. Determine ontologies for describing microarray experiments and the samples used with microarrays (available in RDF, OWL and DAML).
  - Other Ontologies used in GEDs are Taxonomic and Gene Ontologies.
- *MAGE*. Formulates the object model (MAGE-OM), exchange language (MAGE-ML) and software modules (MAGE-stk) for implementing microarray software.
- *Transformations*. Determines recommendations of describing methods for transformations, normalizations and standardizations of microarray data.

Minimum Information Models

CIMR Core Information for Metabolomics Reporting
MIABE Minimal Information About a Bioactive Entity
MIACA Minimal Information About a Cellular Assay
MIAME Minimum Information About a Microarray Experim
MIAME/Env MIAME / Environmental transcriptomic exper
MIAME/Nutr MIAME / Nutrigenomics
MIAME/Plant MIAME / Plant transcriptomics
MIAME/Tox MIAME / Toxicogenomics
MIAPA Minimum Information About a Phylogenetic Analysis
MIAPAR Minimum Information About a Protein Affinity Rea
MIAPE Minimum Information About a Proteomics Experime
MIARE Minimum Information About a RNAi Experiment
MIASE Minimum Information About a Simulation Experiment
MIENS Minimum Information about an ENvironmental Sequence
MIFlowCyt Minimum Information for a Flow Cytometry Experiment
MIGen Minimum Information about a Genotyping Experiment
MIGS Minimum Information about a Genome Sequence
MIMIx Minimum Information about a Molecular Interaction Experiment
MIMPP Minimal Information for Mouse Phenotyping Procedures
MINI Minimum Information about a Neuroscience Investigation
MINIMESS Minimal Metagenome Sequence Analysis Standard
MINSEQE Minimum Information about a high-throughput SeQuencing Experiment
MIPFE Minimal Information for Protein Functional Evaluation
MIQAS Minimal Information for QTLs and Association Studies
MIqPCR Minimum Information about a quantitative Polymerase Chain Reaction experiment
MIRIAM Minimal Information Required In the Annotation of biochemical Models
MISFISHIE Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments
STRENDA Standards for Reporting Enzymology Data
TBC Tox Biology Checklist

BioPAX : Biological Pathways Exchange http://www.biopax.org/
FuGE Functional Genomics Experiment
MGED: Microarray Experimental Conditions

http://www.mibbi.org/index.php/MIBBI_portal



ISA-TAB

- Relating data and its experimental context
  - Investigation, Study, Assay
- TAB = tabular
  - A format suitable for spreadsheets

ISA Defined

- Investigation: high level description of the area and the main aims of a project
- Study: a particular biological hypothesis or analysis
- Assay: specific, individual experiments required to be undertaken together in order to address the study hypotheses

Problems with data

Missing or only partial information:

- Incomplete reactions (products not mentioned)
- Assay conditions missing or reference to another paper
- Kinetic law equation (or fitting equation) not described

Kinetic law types:

no controlled vocabulary used in publications (or even available, except SBO)

→ varying notations referring to several kinetic theories

Parameter units:

- Multiple definitions (e.g. *Katal* or *Unit* for enzyme activities)
- Wrong parameter unit (e.g. *1/s* for Vmax)

Identification of compounds, reactions and enzymes:

- Fuzziness of definitions, conflicting names, wrong synonyms
- Isoenzyme not specified in publication

Gene	Reaction	*E. coli*
Pgi	G6P = F6P	EC4025
Pfk	F6P + ATP − ADP + FP2	EC3916
Fbp	FP2 - F6P + Pi	EC4232
		EC2465
Tal	GAP + Sed7P = Ery4P + F6P	EC2464
TktII	Xyl5P + Ery4P = F6P + GAP	EC2935
		EC2465
Prs	R5P − R5Pex	EC4383

[a]Abbreviations of enzymes: Eno, enolase; Fba, fructose 1,6-bisphosphate aldolase; Fbp, fructose 1,6-bisphosphatase; Gap, glyceraldehyde 3-phosphate dehydrogenase; Gnd, phosphogluconate dehydrogenase (decarboxylating); Gpm, phosphoglycerate mutase; Pfk, 6-phosphofructokinase; Pgi, phosphoglucoisomerase; Pgk, phosphoglycerate kinase; Pgl, phosphogluconolactonase; Ppd, pyruvate, orthophosphate dikinase; Pps, pyruvate, water dikinase (phosphoenolpyruvate synthase); Pyk, pyruvate kinase; Rpe, ribulose-phosphate 3-epimerase; Rpi, ribose 5-phosphate isomerase; Tal, transaldolase; TktI and II, two functions of transketolase; TpiA, triosephosphate isomerase; Zwf, glucose 6-phosphate dehydrogenase. The symbol Prs_DeoB stands for the enzymes catalysing the first step in the conversion of R5P to "external" ribo- and deoxyribonucleotides. When specifying these enzymes to be 5-phosphoribosyl-1-pyrophosphate synthetase (Prs), phosphopentomutase (DeoB) or others, one has to consider the possible additional consumption of ATP. Several enzymes have isoenzymes with different genome identifiers; these are shown, for illustration, only for Pfk and Tkt.

Data integration problems

e.g. Parameter units:

PAR_NAME	PAR_TYPE	START VALUE	END VALUE	UNIT
E	enzymatic activity	0,01	-	nmol*min^(-1)*mg^(-1)
E	enzymatic activity	0,0005	-	nmol/(min*mg of protein)
E	enzymatic activity	0,21	-	nmol/(min*mg)
E	enzymatic activity	2	5,6	units/mg protein
E	enzymatic activity	2	10	units/mg
E	enzymatic activity	1	10	katal

UNIT
nmol*min^(-1)*mg^(-1)
nmol/(min*mg of protein)
nmol/(min*mg)

=nmol/(min*mg)

units/mg protein
units/mg

=U/mg

1 U = the amount of enzyme which catalyses the transformation of 1 µmol of the substrate per minute under standard conditions

Curation

Curation process
(search for errors and inconsistencies)

- *Manually by biological experts*
- *Semi-automatically by consistency checks*
- *Standardisation*
- *Unification*
- *Annotation to controlled vocabularies*
- *Annotation to external datasources*

Annotations of entities in SABIO-RK

Annotations shown to the user:

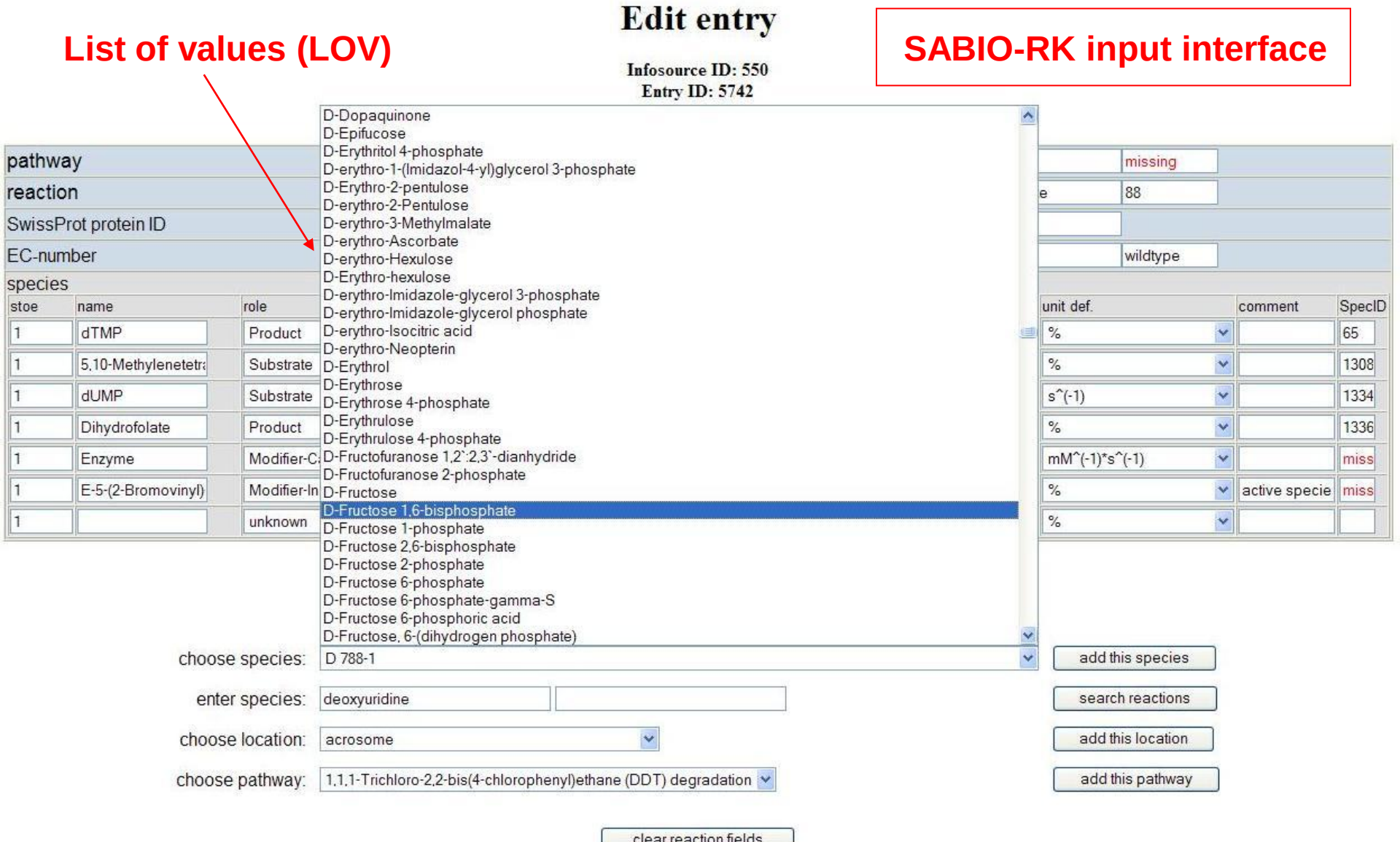
- Chemical compounds to KEGG compound and ChEBI
- Enzymatic activities to Expasy, KEGG, IntEnz, IUBMB and Reactome (query links in the user interface based on the enzyme classification EC)
- Enzyme protein complexes to UniProt/Swiss-Prot
- Cellular locations (compartments etc.) to Gene Ontology (as query link)
- Publications (data sources) to PubMed

Annotations integrated in SABIO-RK, not yet implemented for the output:

- Organisms to NCBI taxonomy
- Kinetic law types and parameter types to SBO (Systems Biology Ontology)
- Species role (substrate, product, modifier, etc.) to SBO
- Reactions to KEGG reactions

More annotations following the MIRIAM standard

Controlled Vocabularies in SABIO-RK



Controlled Vocabularies and Ontologies

Some Biomedical Ontologies:

- **ChEBI (Chemical Entities of Biological Interest):** dictionary and ontological classification of molecular entities focused on ‘small’ chemical compounds
- **Gene Ontology (GO):** cotrolled vocabularies for molecular functions, biological processes and cellular components of gene products
- **Systems Biology Ontology (SBO):** controlled vocabularies and ontologies for systems biology, especially in the context of computational modeling
- **NCBI taxonomy:** controlled vocabulary and classification of organisms

Annotations

Entity	Data Type	URI	Identifier	Qualifier
Enzyme	PubMed	http://www.pubmed.gov	16333295	isDescribedBy
Protein	UniProt	http://www.uniprot.org	P32494	is
Complex	UniProt	http://www.uniprot.org	P32494	hasPart
Reaction	EC class	http://www.ec-code.org	1.1.1.1	isVersionOf

No URL!!

The identifiers for each resource must be unmodifiable/perennial
(not a name that can change like entry names or synonyms)

What is BioPortal

- http://bioportal.bioontology.org/
- Repository for submitting and sharing **Biological ontologies**
- Search for **concepts** across all or selected ontologies
- Tools to link and snip ontologies to create **Views**
- Map concepts with similar meanings across multiple ontologies

Standards & Ontologies for Modeling

Many representations for the same model.

=> Each modeler uses its own representation

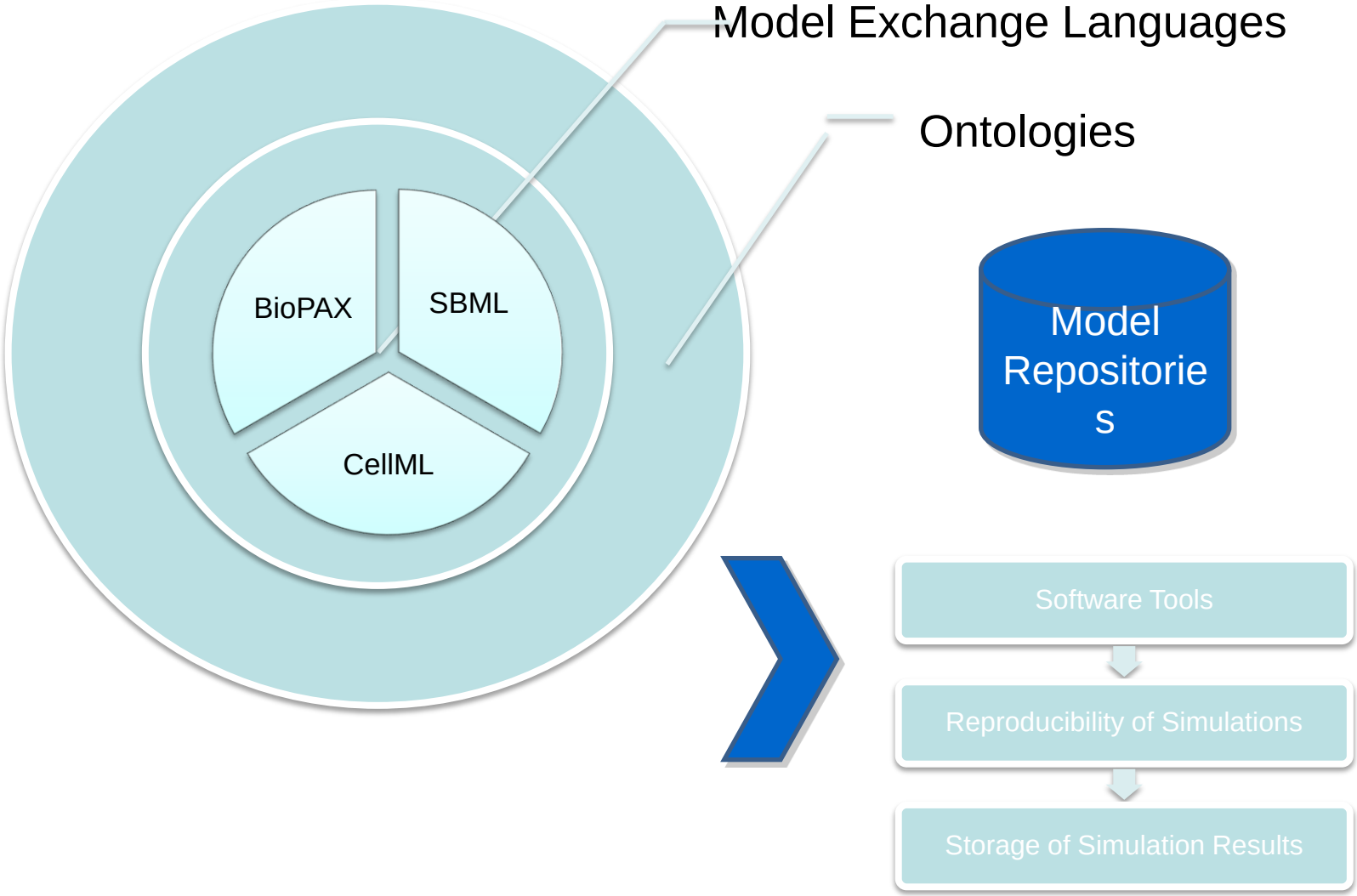
=> Need to read the paper to understand the model

=> Before being able to use or translate models into mathematics, need to understand the modeler's symbols

=> Need for a standard format to be able to

- use models
- exchange models
- compare models
- compose models

Model Exchange Standards

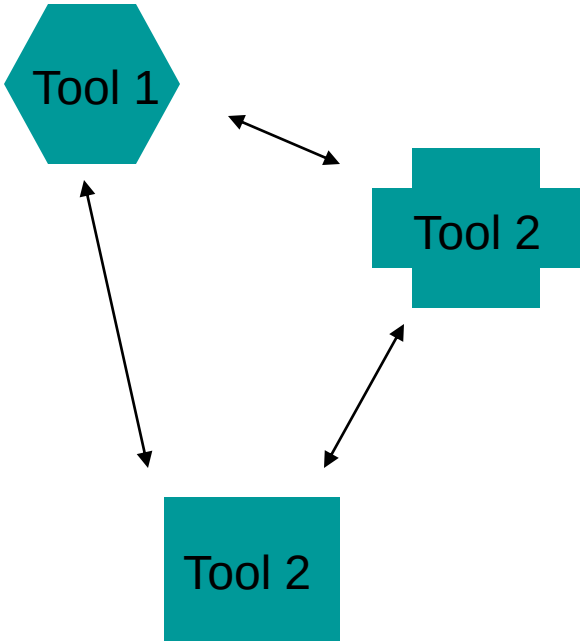


Cell Markup Language

- Designed to support the definition and sharing of models of biological processes.
- Intended to provide consistency in the mathematical representation.
- Encourages model evolution and reuse.
- Started 1999, around same time as SBML.
- CellML and SBML have different scopes:
  - "SBML is designed for representing models of biochemical reaction networks". *(http://www.sbml.org/)*
  - "The purpose of CellML is to store and exchange computer-based mathematical models". *(http://www.cellml.org/)*

[Peter Hunter, *Distributing and maintaining models in CellML,* Life Sciences 2008, Montreal, August, 2008]

SBML Systems Biology Markup Language



The Systems Biology Markup Language (SBML) is a computer-readable format for representing **models of biochemical reaction networks**. SBML is applicable to metabolic networks, cell-signaling pathways, genomic regulatory networks, and many other areas in systems biology.

Originally developed Hamid Bolouri, Andrew Finney, Mike Huck and Herbert Sauro

Structure of SBML

- Beginning of SBML model definition
  - List of function definitions
  - List of unit definitions
  - List of compartment types
  - List of molecular species types
  - List of compartments
  - List of species
  - List of parameters
  - List of initial assignments
  - List of rules
  - List of constraints
  - List of reactions
  - List of events
- End of SBML model definition

Systems Biology Markup Language

- Broad Acceptance
  - Supported by over 100 software systems
    - Simulators
    - Databases
    - Analysis tools
    - Editing tools
  - Supported by several alliances

    DARPA Bio-SPICE, IECA, others
  - Supported by journals

    "*Nature* journals and *Molecular Systems Biology* support submissions involving SBML." [*Nature*, p.1, May 5, 2005]

Ontologies

- MIRIAM – Minimal Information Requested In the Annotation of biochemical Models
- SBO – Systems Biology Ontology
- TEDDY - TErminology for the Description of Dynamics

- In Conception / Planning
  - KiSAO - Kinetic Simulation Algorithm Ontology

Proposed Standard: MIRIAM

Minimum **I**nformation **R**equested **I**n the **A**nnotation of biochemical **M**odels

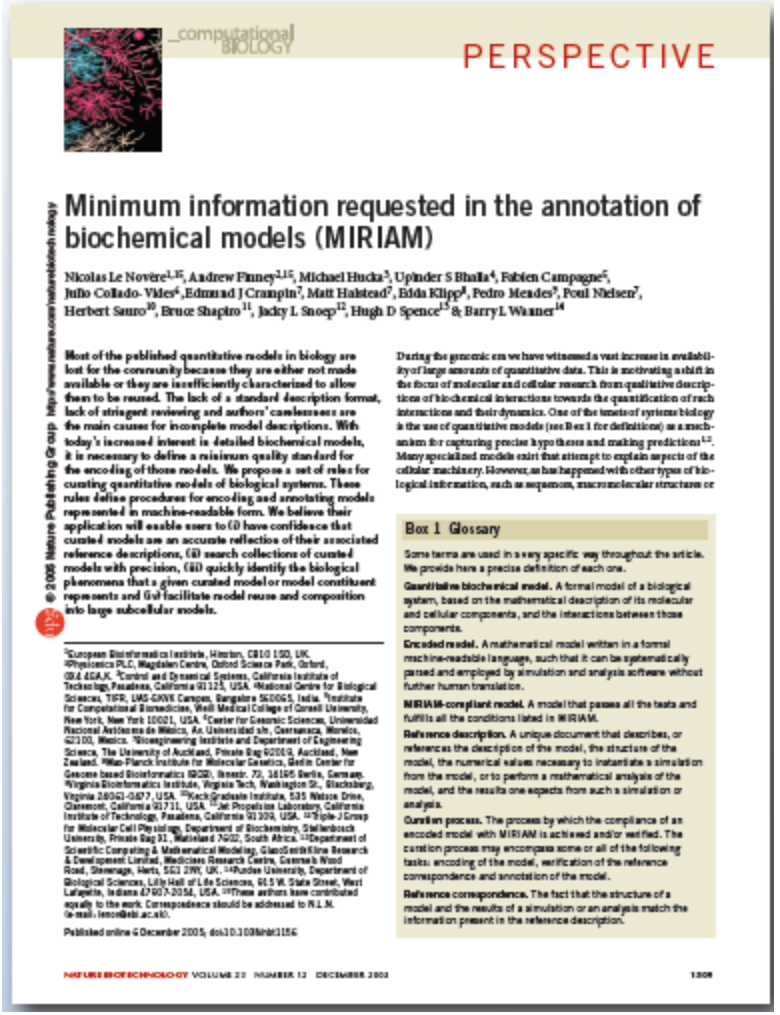
Proposed guidelines for annotation and curation of **quantitative** models

- Specifically about encoding & annotation
- Limited to models that can be simulated

MIRIAM approach avoids putting *data content* directly into the model;

instead, it points at external resources that contain the knowledge.

www.ebi.ac.uk/miriam



computational BIOLOGY

PERSPECTIVE

Minimum information requested in the annotation of biochemical models (MIRIAM)

Nicolas Le Novère[1,15], Andrew Finney[2,15], Michael Hucka[3], Upinder S Bhalla[4], Fabien Campagne[5], Julio Collado-Vides[6], Edmund J Crampin[7], Matt Halstead[7], Edda Klipp[8], Pedro Mendes[9], Poul Nielsen[7], Herbert Sauro[10], Bruce Shapiro[11], Jacky L Snoep[12], Hugh D Spence[13] & Barry L Wanner[14]

Most of the published quantitative models in biology are lost for the community because they are either not made available or they are inefficiently characterized to allow them to be reused. The lack of a standard description format, lack of stringent reviewing and authors' carelessness are the main causes for incomplete model descriptions. With today's increased interest in detailed biochemical models, it is necessary to define a minimum quality standard for the encoding of those models. We propose a set of rules for curating quantitative models of biological systems. These rules define procedures for encoding and annotating models represented in machine-readable form. We believe their application will enable users to (i) have confidence that curated models are an accurate reflection of their associated reference descriptions, (ii) search collections of curated models with precision, (iii) quickly identify the biological phenomena that a given curated model or model constituent represents and (iv) facilitate model reuse and composition into large subcellular models.

During the genomic era we have witnessed a vast increase in availability of large amounts of quantitative data. This is motivating a shift in the focus of molecular and cellular research from qualitative descriptions of biochemical interactions towards the quantification of such interactions and their dynamics. One of the tenets of systems biology is the use of quantitative models (see Box 1 for definition) as a mechanism for capturing precise hypotheses and making predictions[1,2]. Many specialized models exist that attempt to explain aspects of the cellular machinery. However, as has happened with other types of biological information, such as sequences, macromolecular structures or

Box 1 Glossary

Some terms are used in a very specific way throughout the article. We provide here a precise definition of each one.

Quantitative biochemical model. A formal model of a biological system, based on the mathematical description of its molecular and cellular components, and the interactions between those components.

Encoded model. A mathematical model written in a formal machine-readable language, such that it can be systematically parsed and employed by simulation and analysis software without further human translation.

MIRIAM-compliant model. A model that passes all the tests and fulfills all the conditions listed in MIRIAM.

Reference description. A unique document that describes, or references the description of the model, the structure of the model, the numerical values necessary to instantiate a simulation from the model, or to perform a mathematical analysis of the model, and the results one expects from such a simulation or analysis.

Curation process. The process by which the compliance of an encoded model with MIRIAM is achieved and/or verified. The curation process may encompass some or all of the following tasks: encoding of the model, verification of the reference correspondence and annotation of the model.

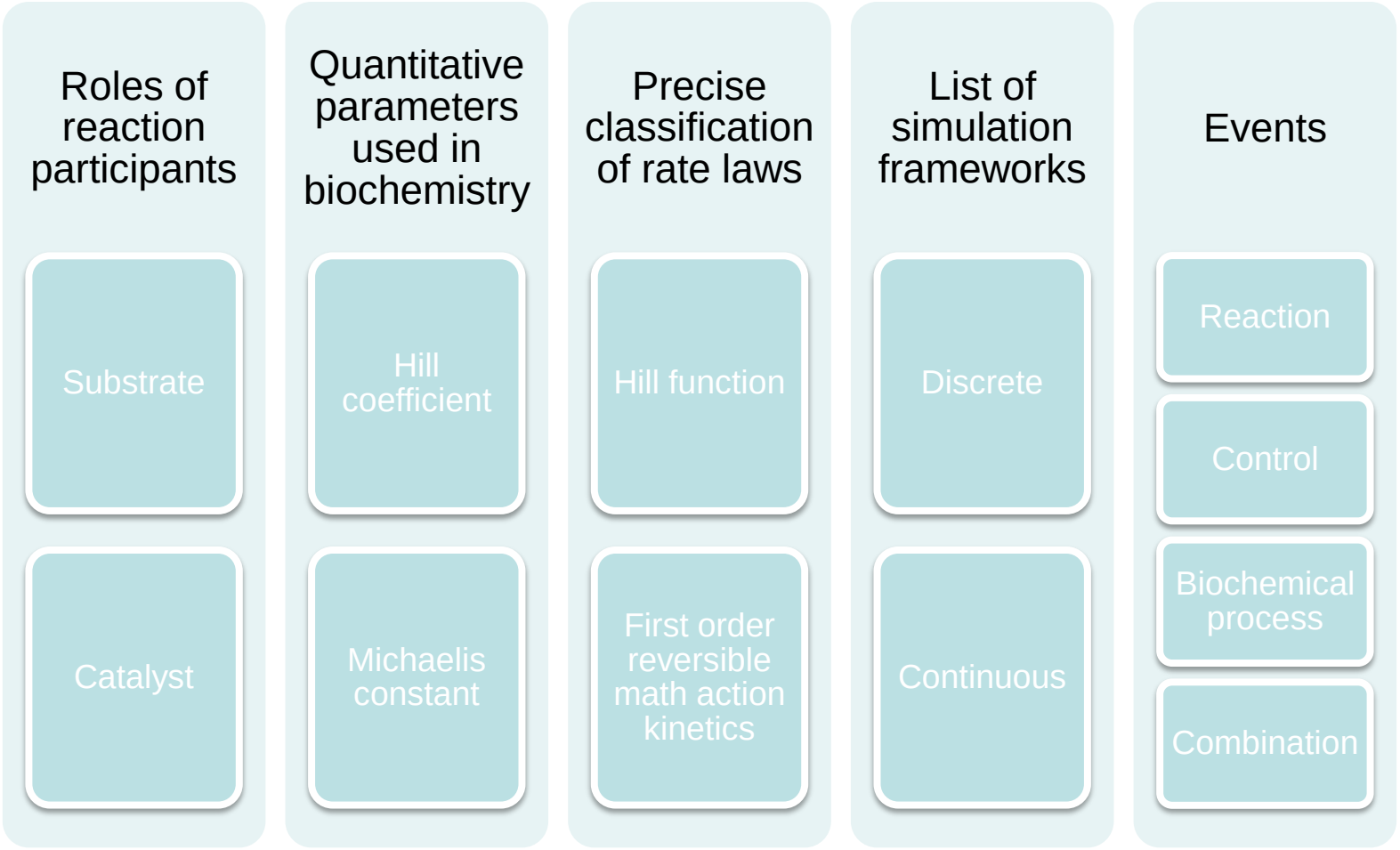
Reference correspondence. The fact that the structure of a model and the results of a simulation or an analysis match the information present in the reference description.

[1]European Bioinformatics Institute, Hinxton, CB10 1SD, UK. [2]Physiomics PLC, Magdalen Centre, Oxford Science Park, Oxford, OX4 4GA, UK. [3]Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA. [4]National Centre for Biological Sciences, TIFR, UAS-GKVK Campus, Bangalore 560065, India. [5]Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, New York 10021, USA. [6]Center for Genomic Sciences, Universidad Nacional Autónoma de México, Av. Universidad s/n, Cuernavaca, Morelos, 62100, Mexico. [7]Bioengineering Institute and Department of Engineering Science, The University of Auckland, Private Bag 92019, Auckland, New Zealand. [8]Max-Planck Institute for Molecular Genetics, Berlin Center for Genome based Bioinformatics (BCB), Ihnestr. 73, 14195 Berlin, Germany. [9]Virginia Bioinformatics Institute, Virginia Tech, Washington St., Blacksburg, Virginia 24061-0477, USA. [10]Keck Graduate Institute, 535 Watson Drive, Claremont, California 91711, USA. [11]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [12]Triple-J Group for Molecular Cell Physiology, Department of Biochemistry, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. [13]Department of Scientific Computing & Mathematical Modeling, GlaxoSmithKline Research & Development Limited, Medicines Research Centre, Gunnels Wood Road, Stevenage, Herts, SG1 2NY, UK. [14]Purdue University, Department of Biological Sciences, Lilly Hall of Life Sciences, 915 W. State Street, West Lafayette, Indiana 47907-2054, USA. [15]These authors have contributed equally to the work. Correspondence should be addressed to N.L.N. (lenov@ebi.ac.uk).

Published online 6 December 2005; doi:10.1038/nbt1156

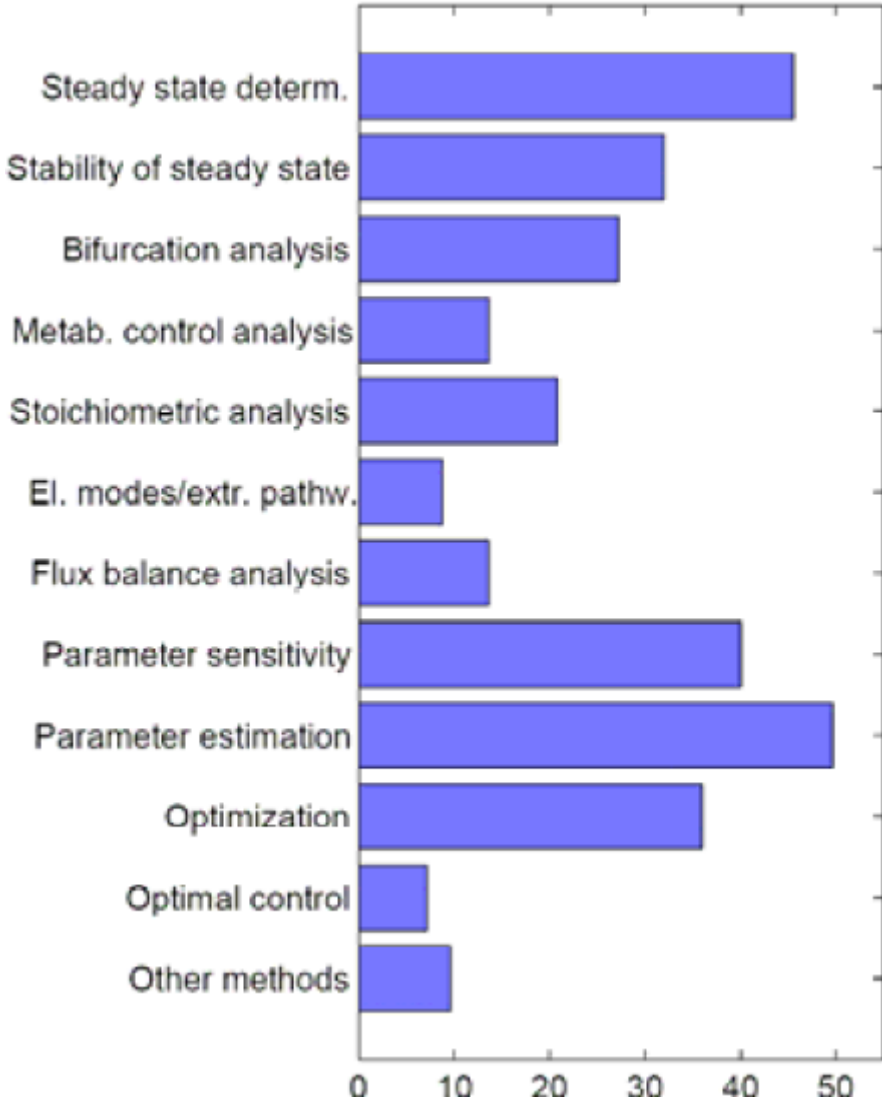
Nature Biotech. 23(12), Dec. 2005

SBO – Controlled Vocabularies

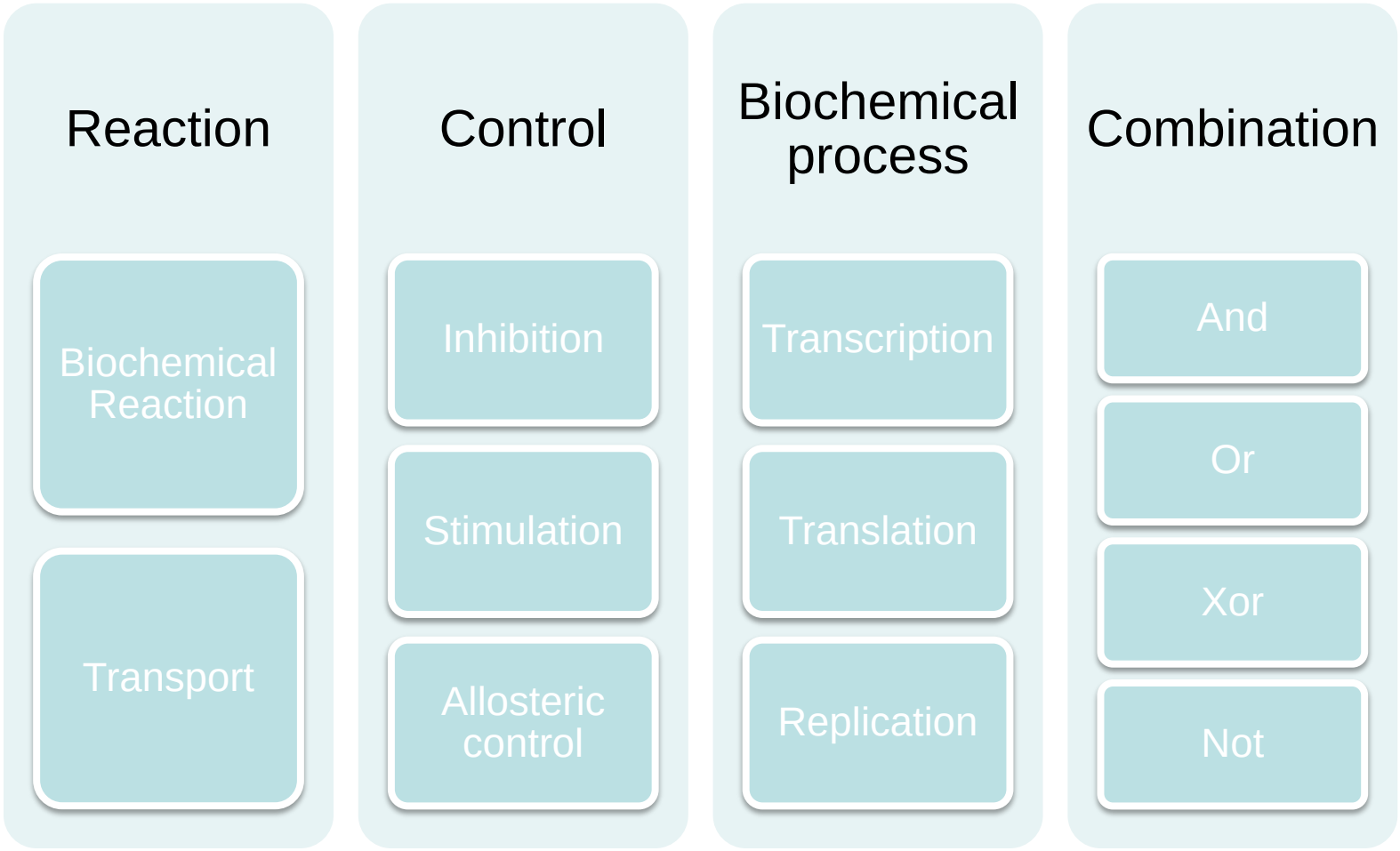


Model types and analysis methods

Deterministic
Non-Deterministic
Probabilistic
Discrete
Continuous
State transition
ODE (Ordinary Differential Equations)
PDE (Partial Differential Equations)
linear equations
Steady state
Stochiometric



Events


Reaction
Biochemical Reaction
Transport
Control
Inhibition
Stimulation
Allosteric control
Biochemical process
Transcription
Translation
Replication
Combination
And
Or
Xor
Not

Reproducing / Evaluating Simulation Results

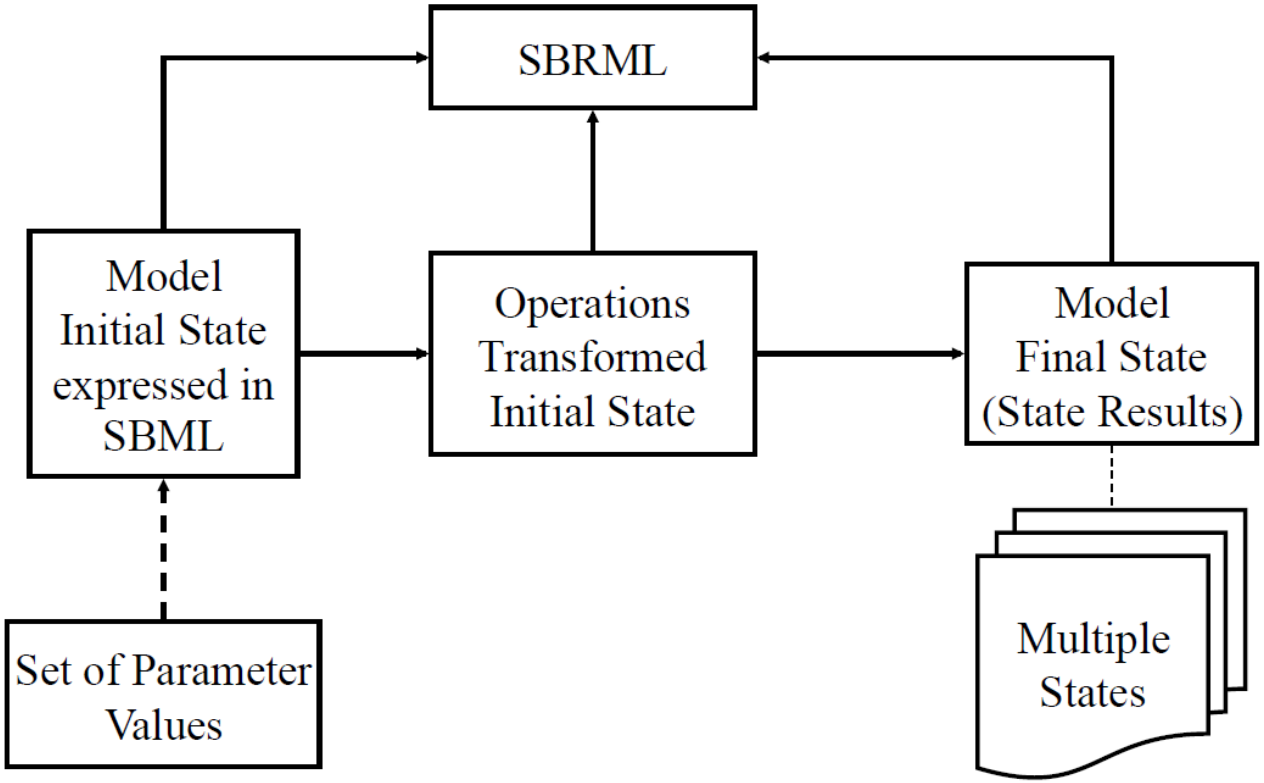
- MIASE - Minimum Information needed About a Simulation Experiment
- SBRML - Systems Biology Results Markup Language

MIASE – Why is it necessary?

- Problem:
  - "The model, when instantiated within a suitable simulation environment, must be able to reproduce all relevant results given in the reference description that can readily be simulated"
- However: MIRIAM does not include guidelines about how a relevant result can be reproduced

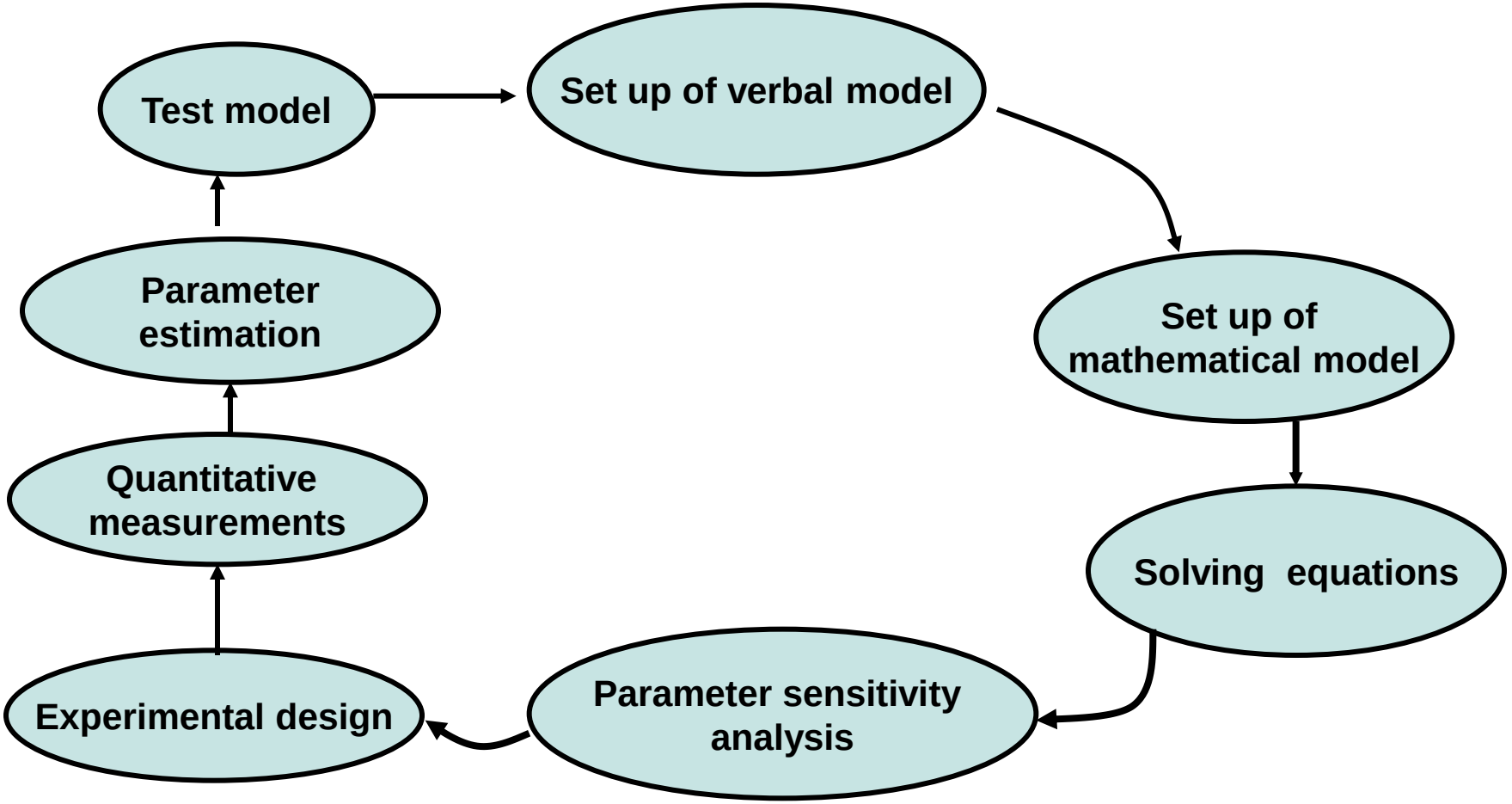
SBRML

- Aim: Linking numerical results to the model that gave rise to them



Bottom-up: modelling cycle

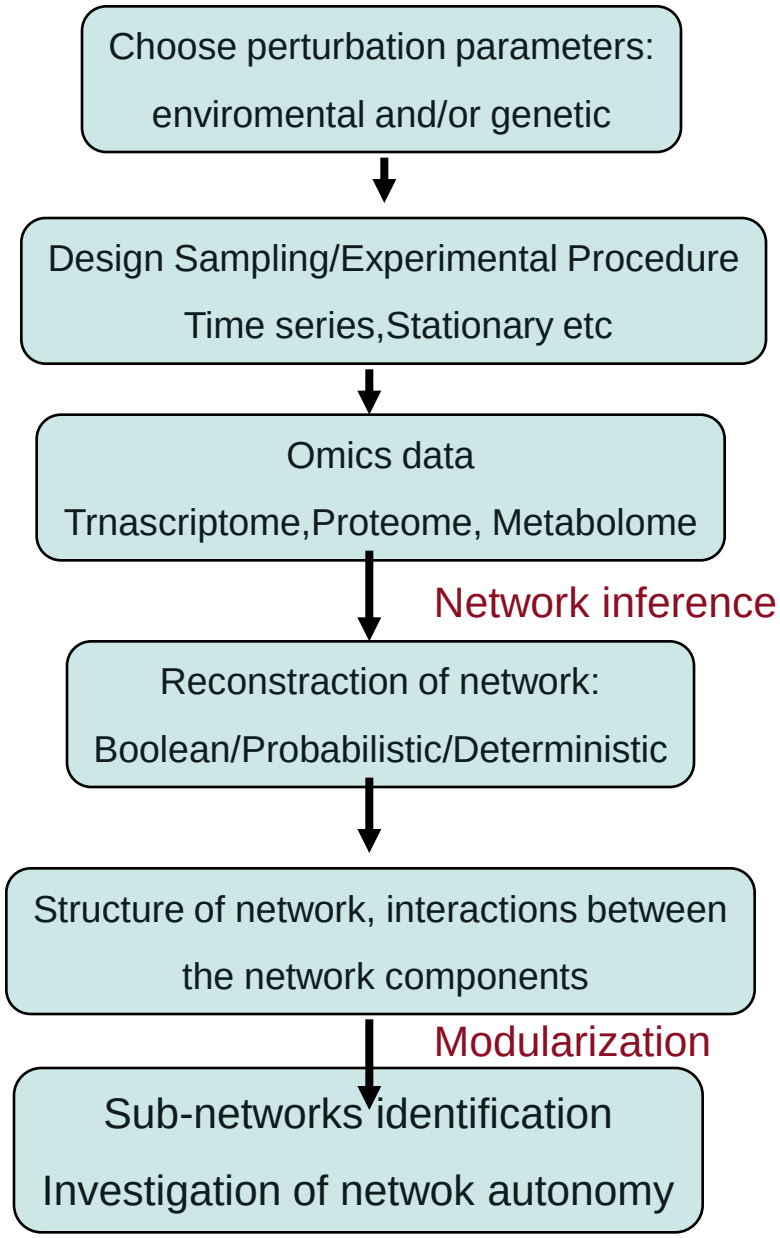
Forward engineering



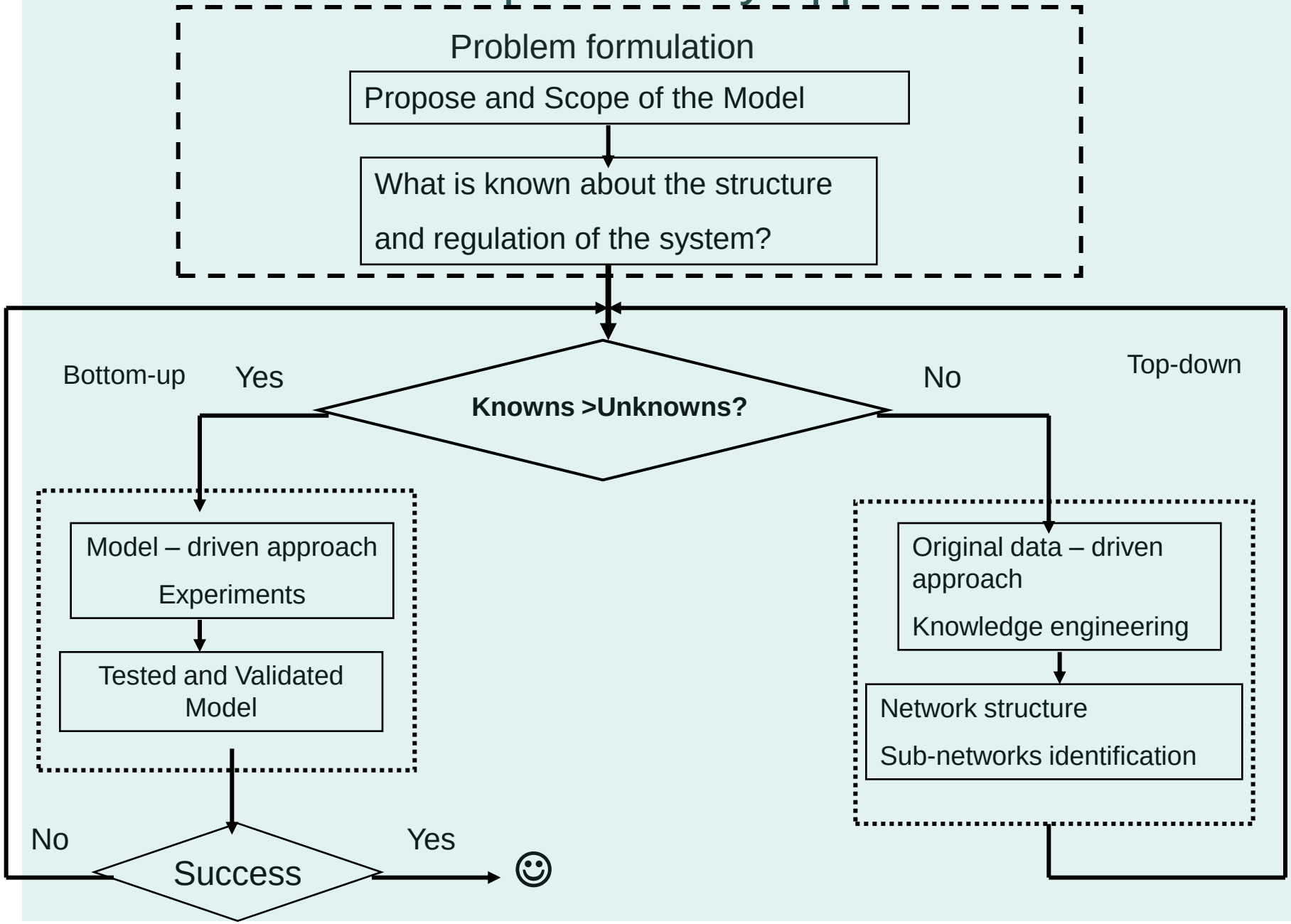
M.Reuss, from “ESF forward look report - Systems Biology: a Grand Challenge for Europe”, 2007

Top-down

Reverse engineering



Mixed complementary approach



From M.Reuss, from “ESF_ Systems Biology: a Grand Challenge for Europe”, 2007

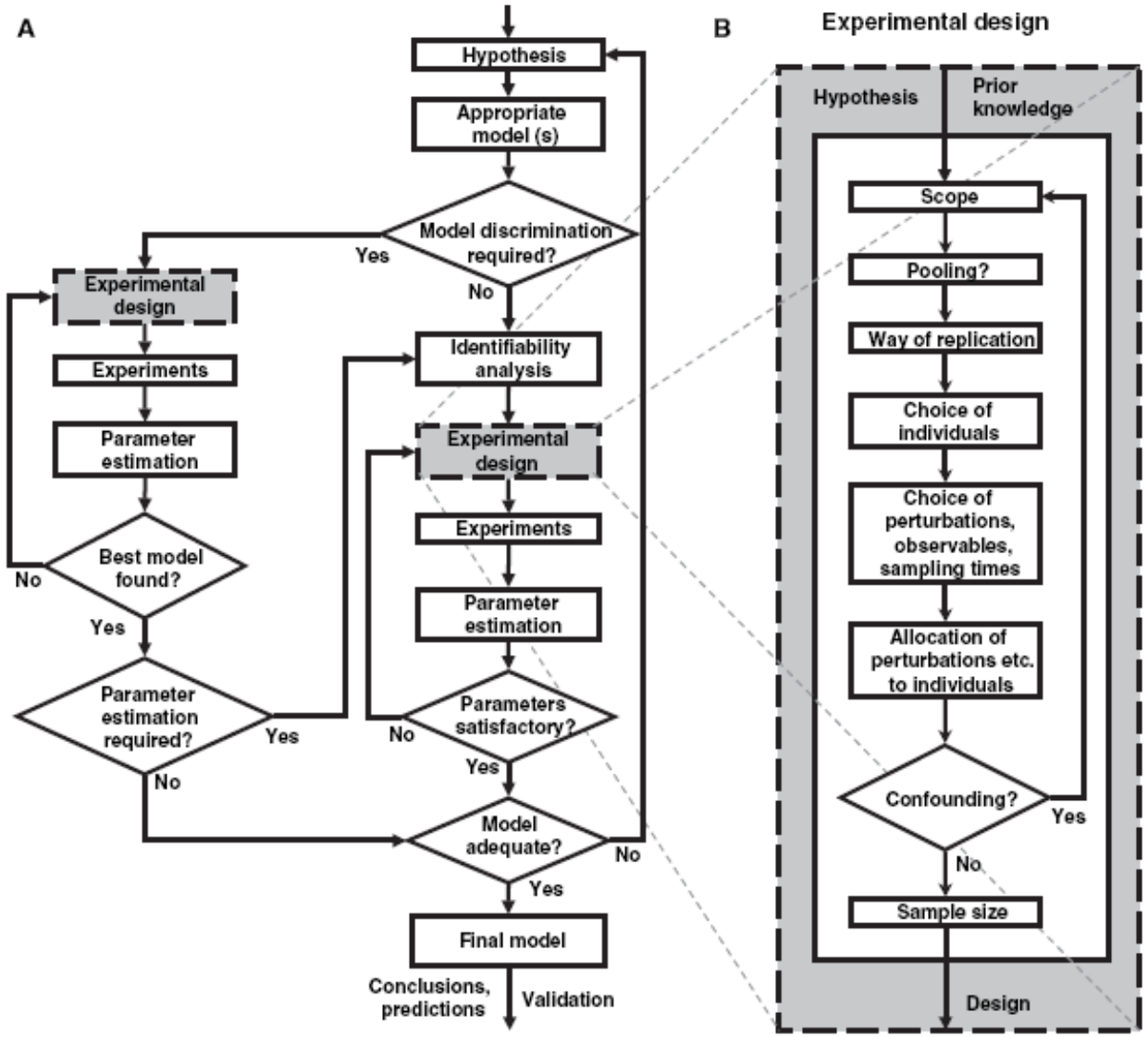
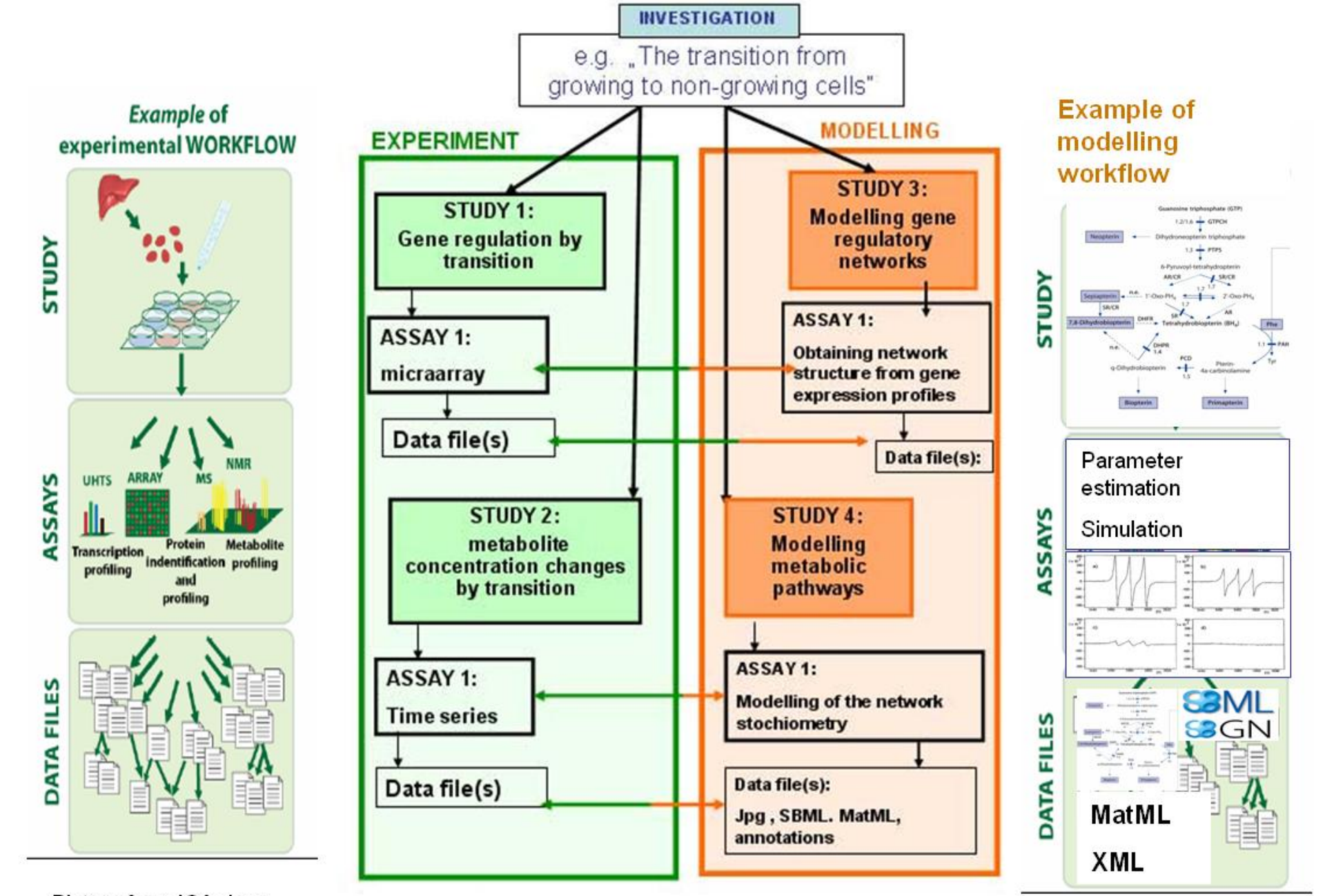


Fig. 1. (A) Overview of an usual model building process. Both loops, with and without model discrimination, require experimental planning (highlighted in gray). (B) The most important steps in experimental planning for systems biological applications.

J.Timmer at al., 2008

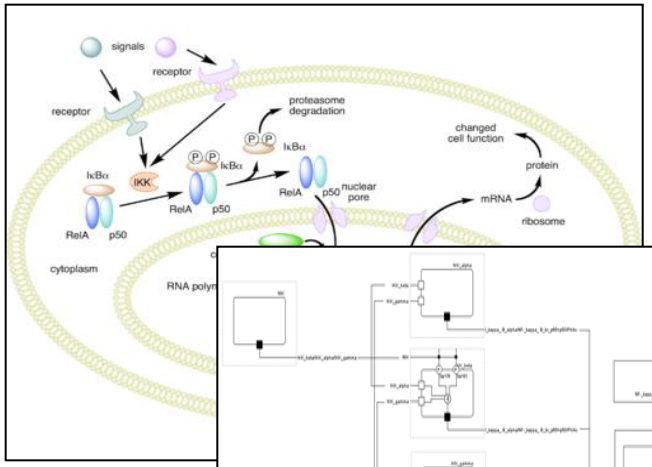


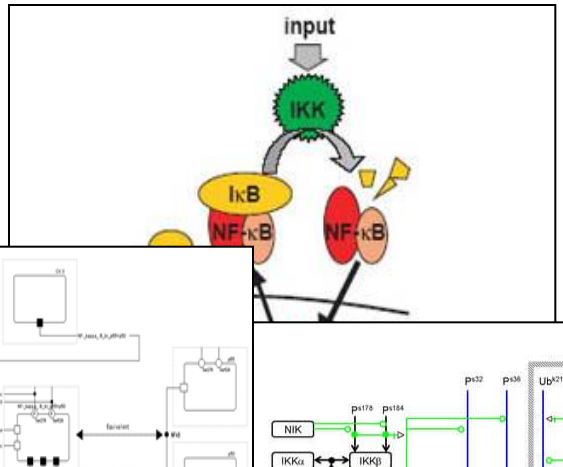
Picture from ISA docu, http://isatab.sourceforge.net/docs/ISA-infrastructure-overview-25June09.pdf

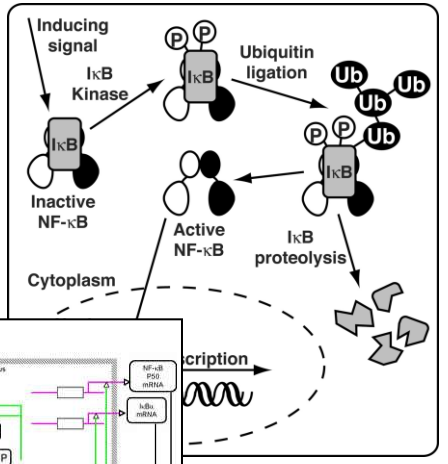
Networks visualisation

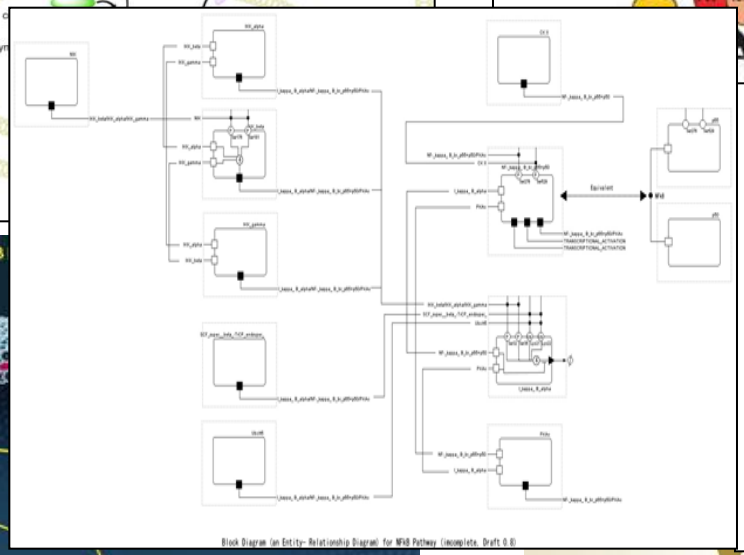
Variety of models: example of NFkB

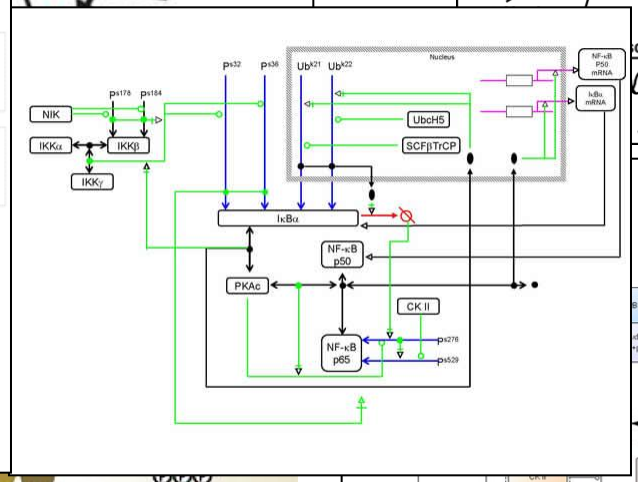
A model can be represented in many different ways:

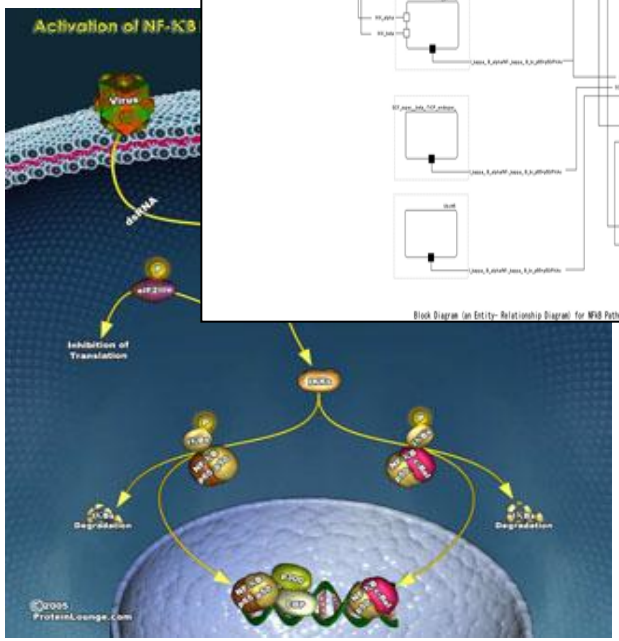


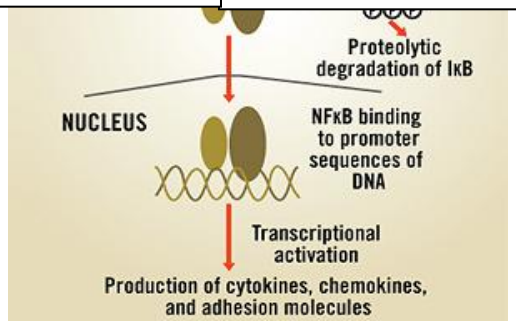


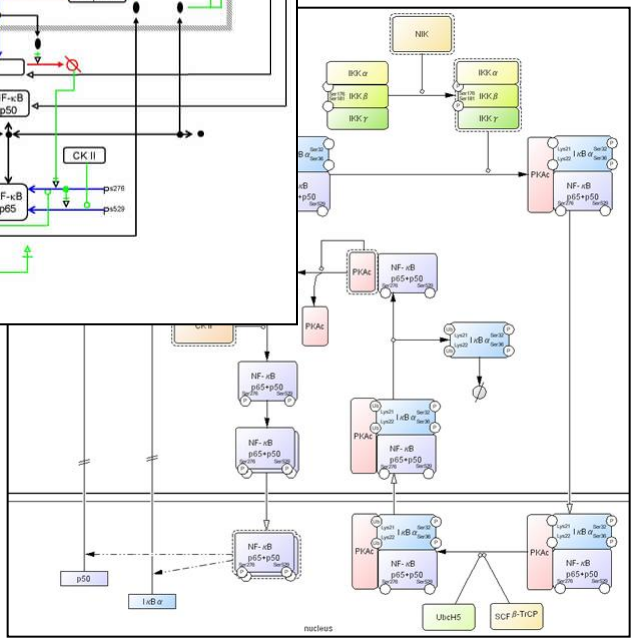




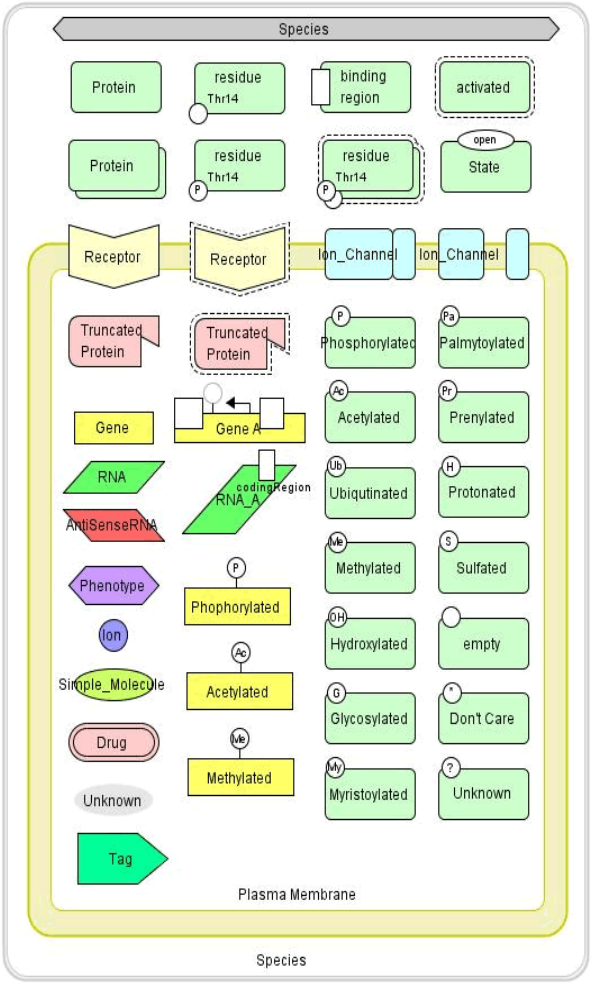


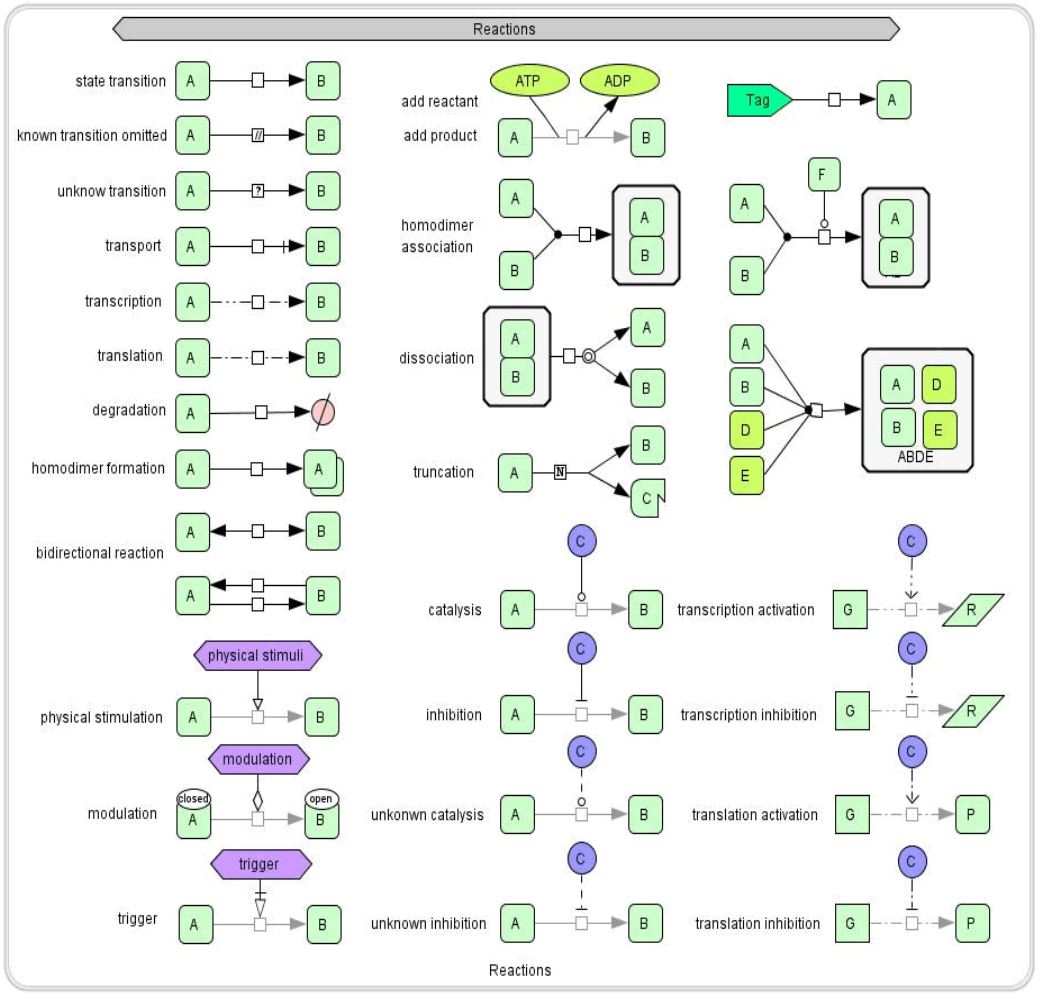


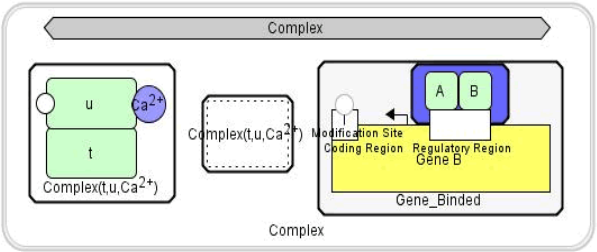


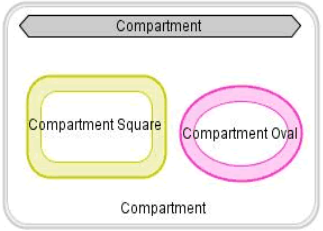


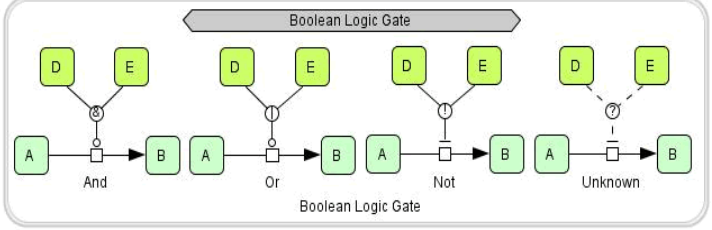
Graphical representation: SBGN - http://sbgn.org/







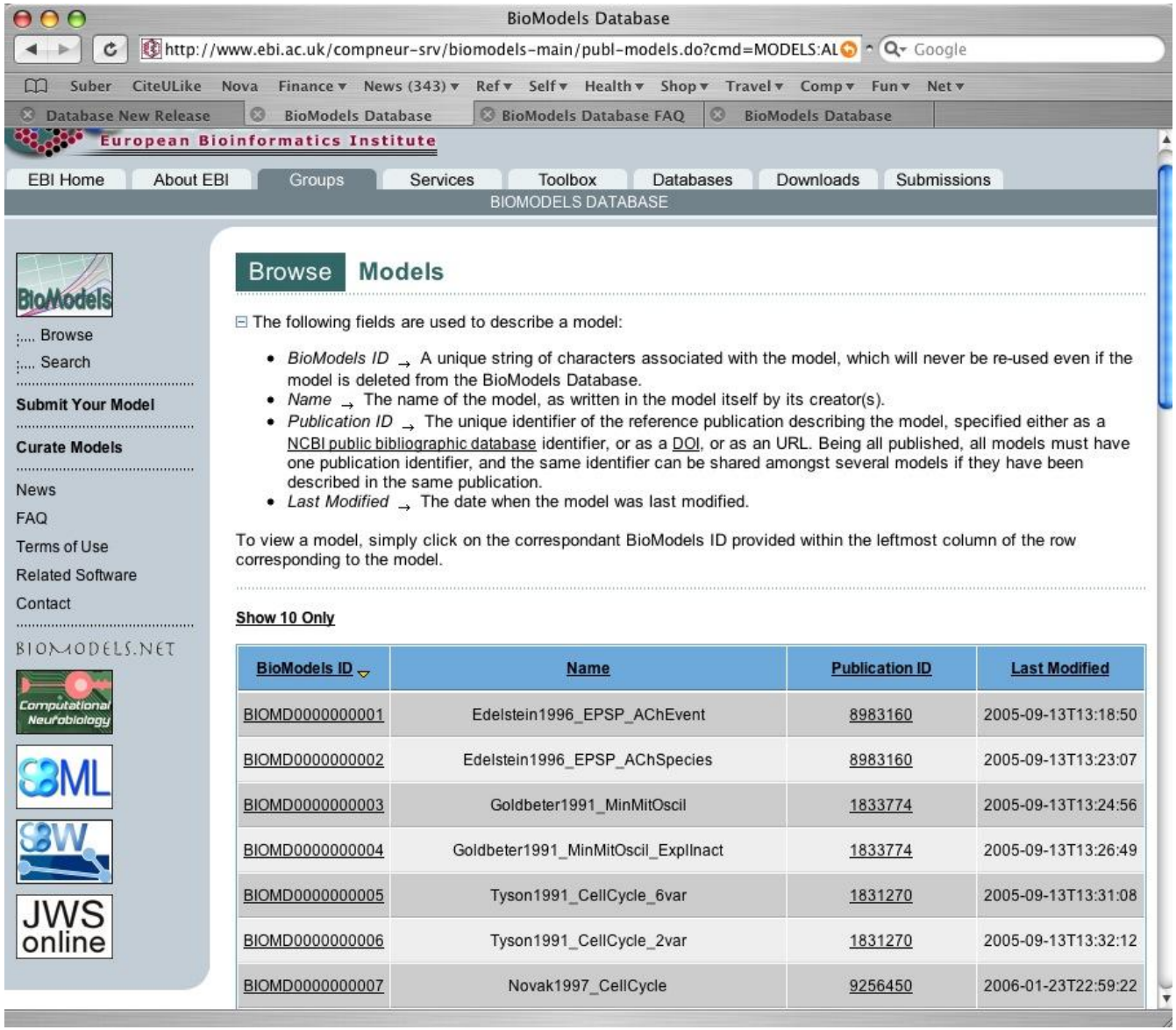




BioModels Database

- Stores & serves quantitative models of biomed. interest
  - Only models described in peer-reviewed scientific literature
- Models are curated by humans: computer software checks syntax, humans check semantics
- Models are simulated to check correspondence to reference
- Model components are annotated to improve identification and retrieval
- Accepted in SBML and CellML formats, served in several (SBML, XPP, CellML, diagram; more coming)

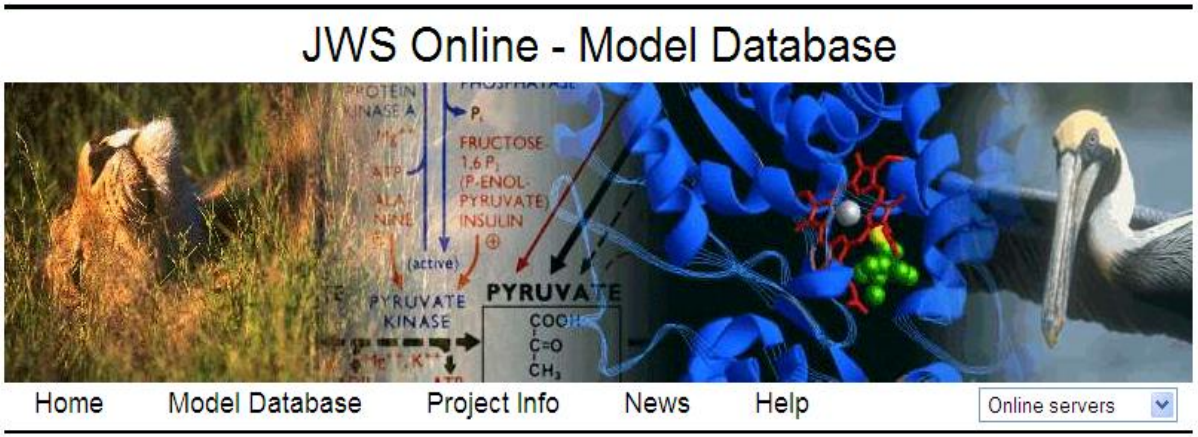
www.ebi.ac.uk/biomodels



CellML Repository

- CellML model repository has over 350 published models of:
  - Signal transduction pathways;
  - Metabolic pathways;
  - Electrophysiological;
  - Calcium dynamics;
  - Immunology;
  - Cell cycle;
  - Smooth and skeletal muscle models;
  - Mechanical and constitutive relationships.
- **http://www.cellml.org/models/**

[Peter Hunter, *Distributing and maintaining models in CellML,* Life Sciences 2008, Montreal, August, 2008]



Publish, manage, run, validate SBML models

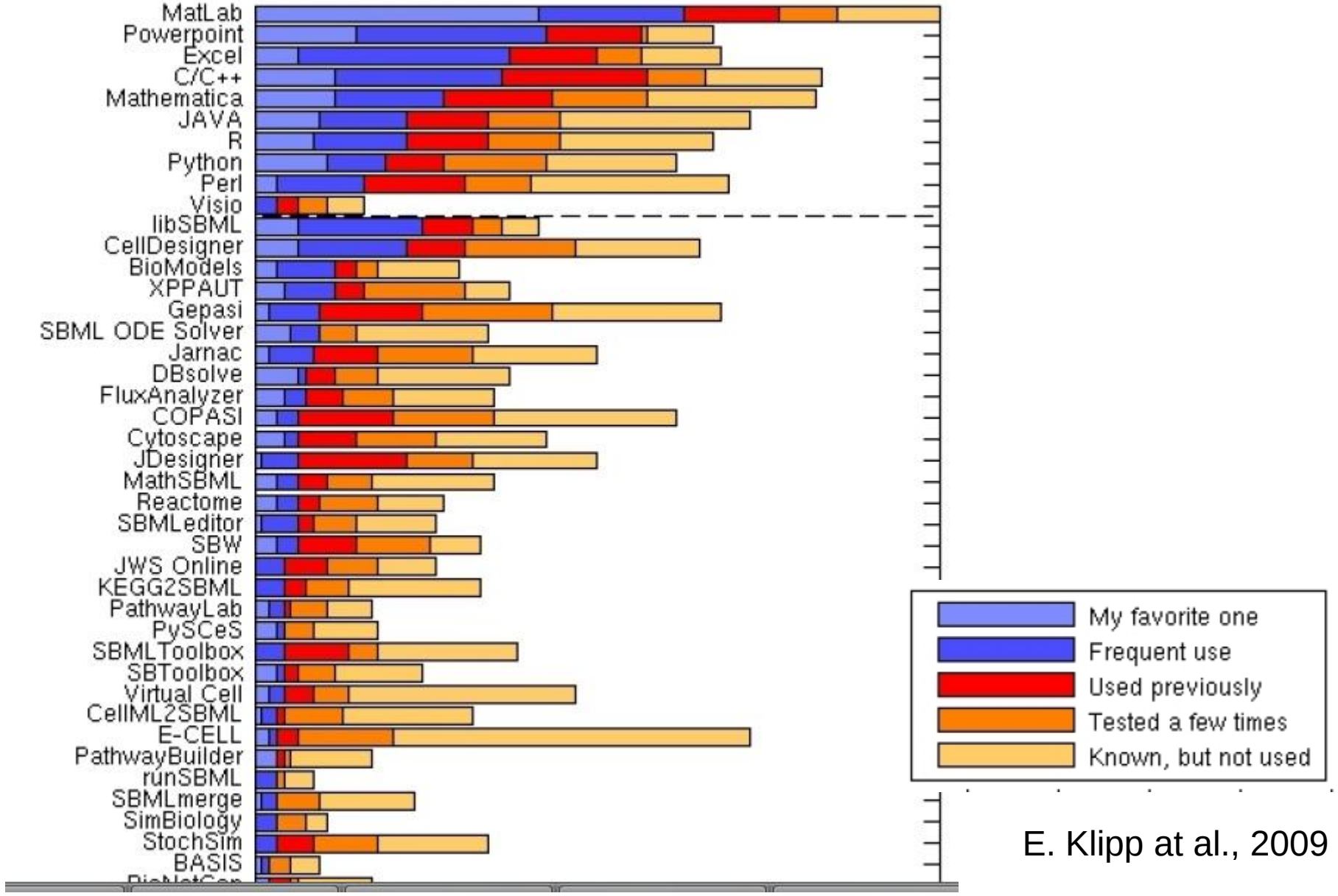
- Database of curated models and a model simulator
- Web service enabled to run from workflows
- Separate password protected websites for each project
- Special instance of JWS Online for SysMO
- Validate and run models
- Access control
- Access to other resources (Biomodels, Copasi)
- Semantic SBML from TRANSLUCENT project
- SBML and MIRIAM

Physiome.org Repository

- The physiome.org model repository has about 300 live models (and is undergoing a big revision):
  - Convection-diffusion reaction and exchange;
  - Tissue and organ models for PET and MRI analysis;
  - Electrophysiology, mostly cardiac;
  - Physico-chemical, osmotic processes, cells, tissues;
  - Enzymatic reactions, metabolic networks;
  - Transport and exchange of respiratory gases, ;
  - Circulatory and respiratory mechanics, coupled with transport and control mechanisms;
  - Membrane transporters and pumps.
  - Pharmacokinetic models for complex systems.
  - Families of tutorials for physiological transport and metabolism.
- **http://www.physiome.org/Models/**

Raymond GM, Butterworth E, and Bassingthwaighte JB. JSIM: Free software package for teaching physiological modeling and research. Exper Biol 2003 280.5, p102, 2003.

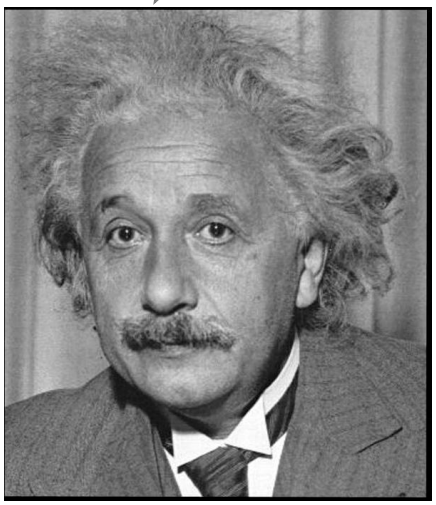
Popularity of software tools



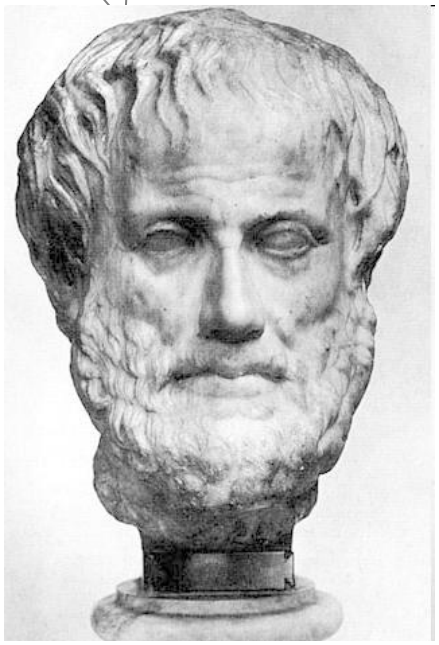
E. Klipp at al., 2009

Conclusions?

Things should be made as simple as possible but not simpler.



The whole is more than the sum of its parts

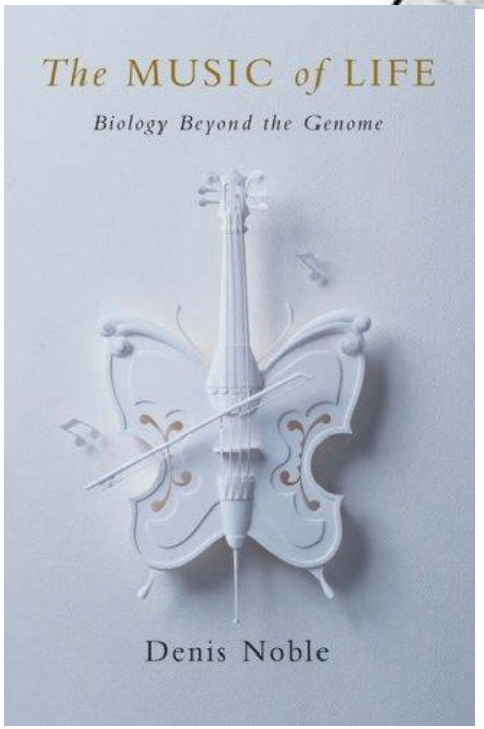


What is life?

So asked the distinguished physicist Erwin Schroedinger in his famous lecture at Trinity College Dublin in 1943. Now, after the full mapping of the human genome has yielded a code of three billion letters, we are still far from a satisfactory answer to this question. Denis Noble

The reductionist approach of molecular biology has proved itself immensely powerful. But DNA isn't life. It doesn't even leave the nucleus of the cell.

We must look not at one level, but at the interaction of processes at various levels, from the realm of systems biology, a field that has been growing in strength in the past decade.



Acknowledgements

- All the hardworking and altruistic colleagues developing standards&ontologies
- HITS/SDBV
- and you for your attention ☺