# Presentation of thesis project

Athénaïs Vaginay

2018 november 28

# Outline

the subject is at the meeting point of

- modelling biological systems
- semantic web technologies

# vertical hierarchy

Many levels in biology $\rightarrow$(Multi-)scale biological models

atom
molecule
macromolecule
**→cell←**
tissue
organ
organ
organism
population
community
ecosystem
biosphere

# horizontal modularity : signaling, gene regulation, metabolism

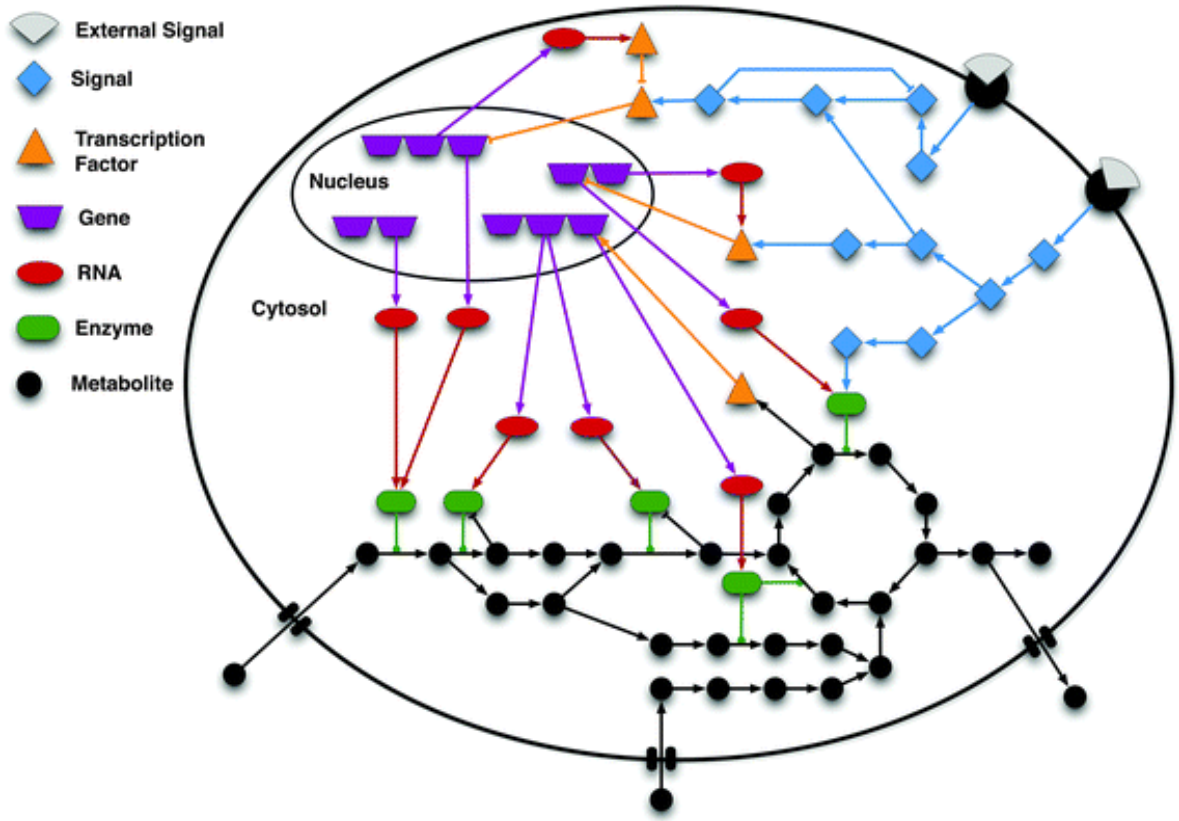
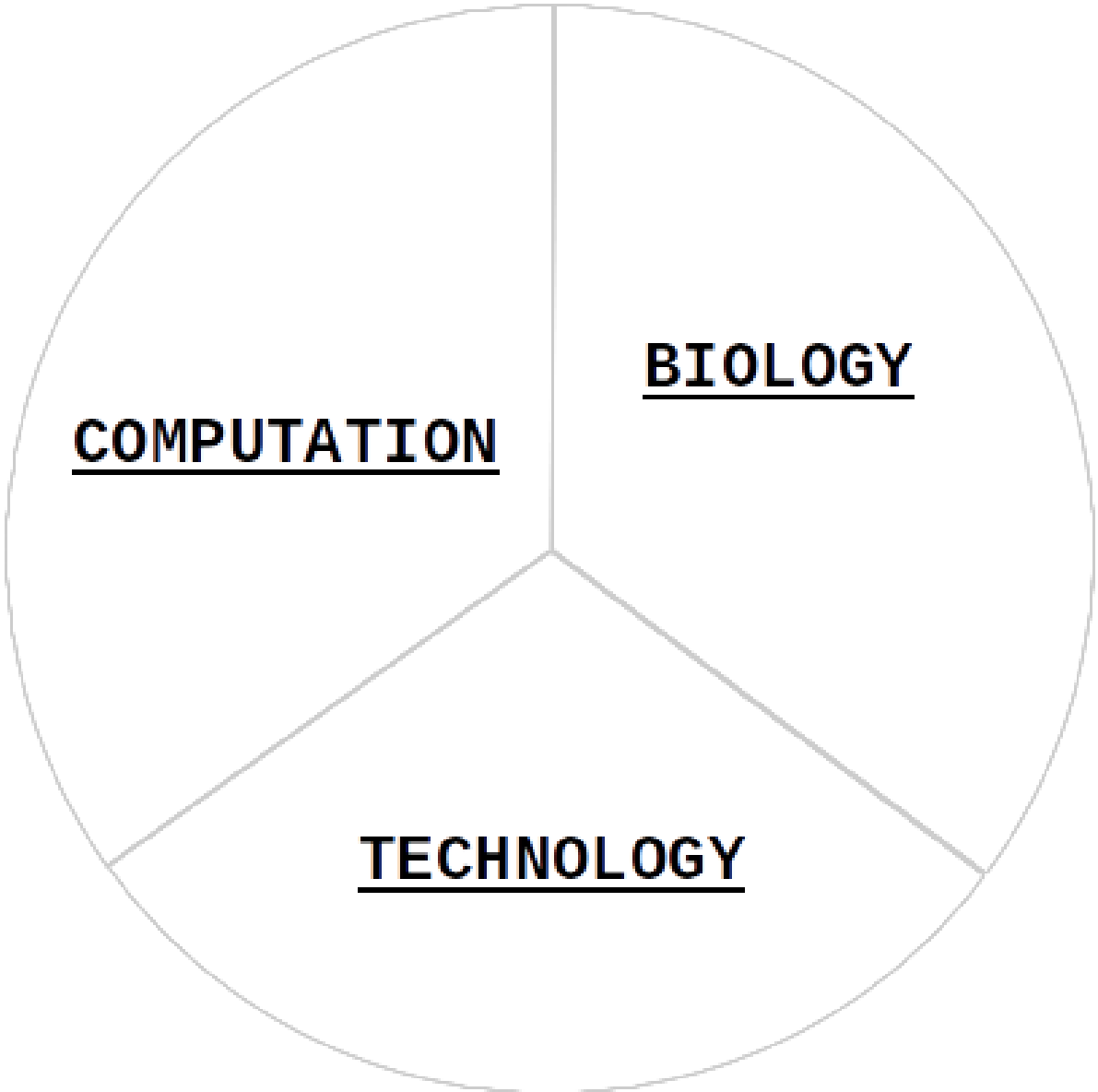
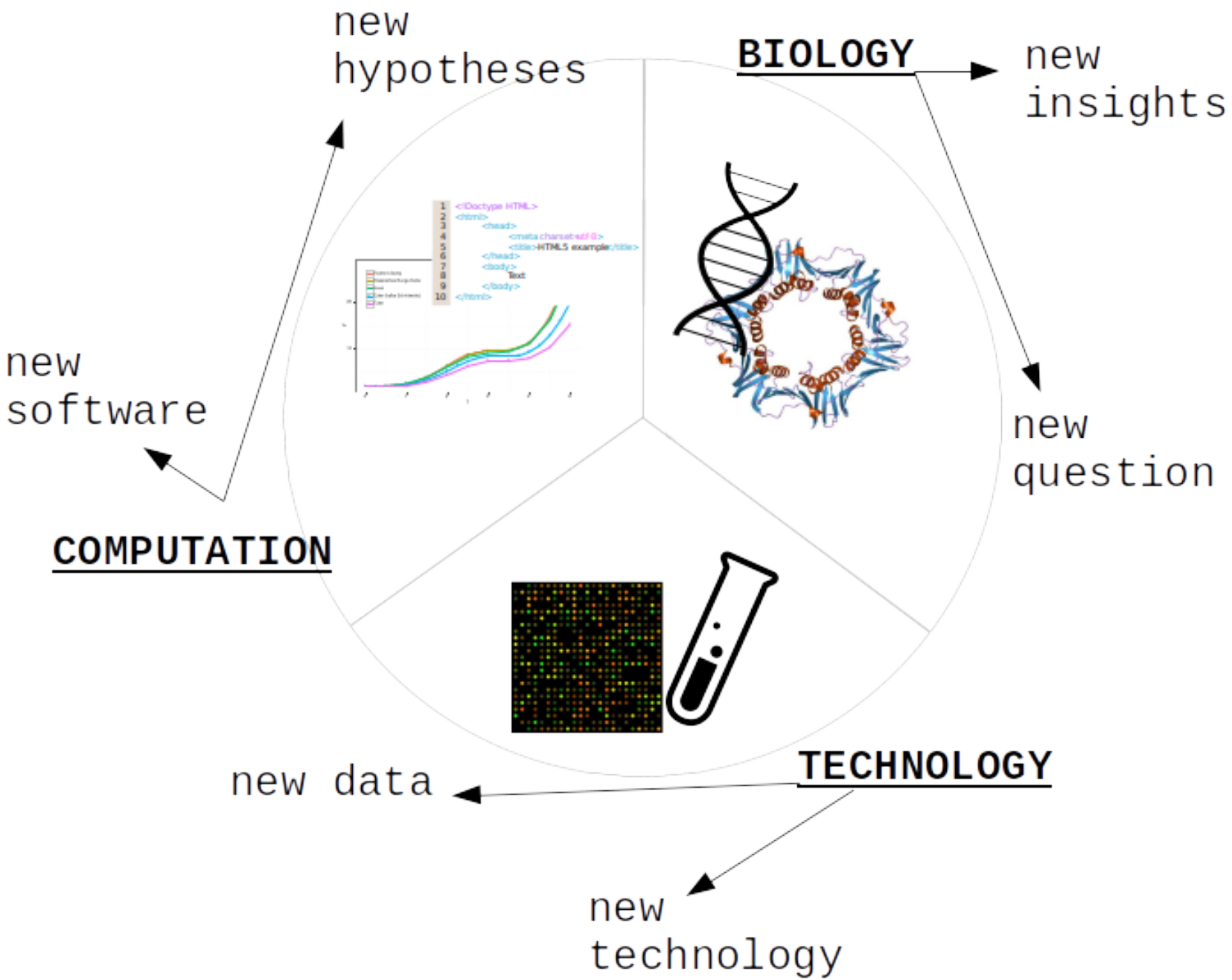



Figure 1: Gonçalves et al. "Bridging the Layers: Towards Integration of Signal Transduction, Regulation and Metabolism into Mathematical Models." Molecular BioSystems 9, no. 7 (2013): 1576. https://doi.org/10.1039/c3mb25489e

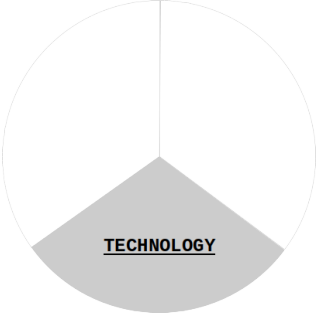
# systems biology triade cycle



# systems biology triade cycle



# data acquisition

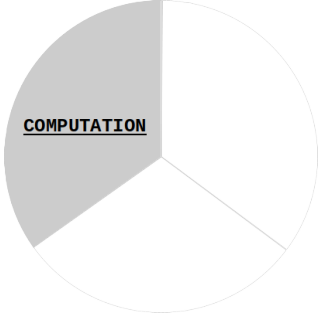



Different kind of data.

- images, kinetic parameters (tedious) eg. SABIO-RK Biochemical Reaction Kinetics Database, omics (dimentionality problem)
- cell population (hide biological variability), single cell (no time series datasets)

Developement of new technologies to retrieve data.

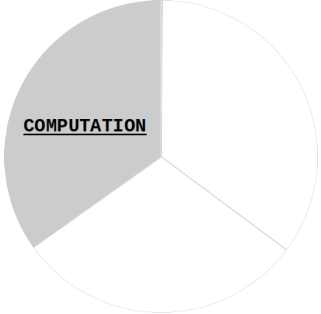
# model generation and analysis



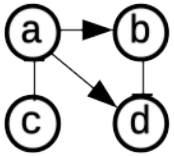

a LOT of formalisms

- mechanistic models
- statistical model

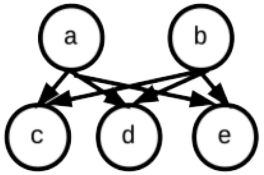
# formalisms - mechanistic models




Boolean network



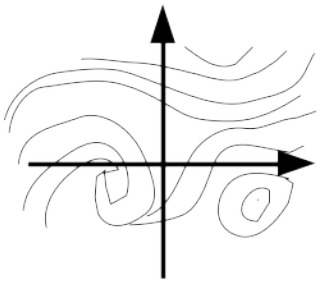

Bayesian network



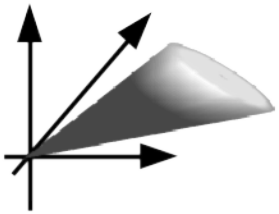

Process algebras

```
((b(x,de)[E]) || (B(y, dI)[I]))
bh(x, dE) bh(y, dI) (E || I)
```

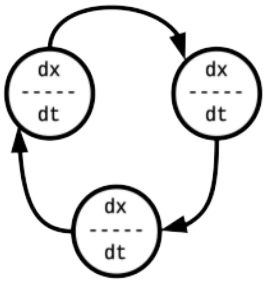
Differential equation



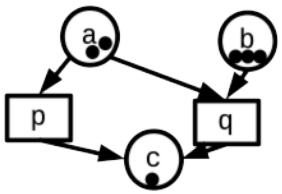
Constraint based model



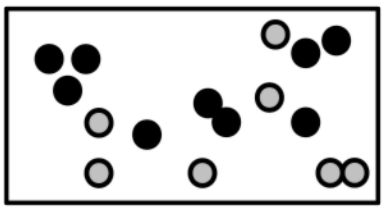
Hybrid systems



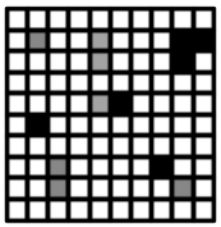

Petri Nets




Agent-based model



Cellular automata



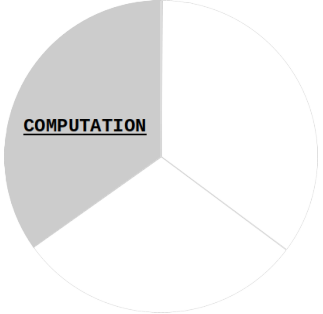
Interacting state machine

Compartment based

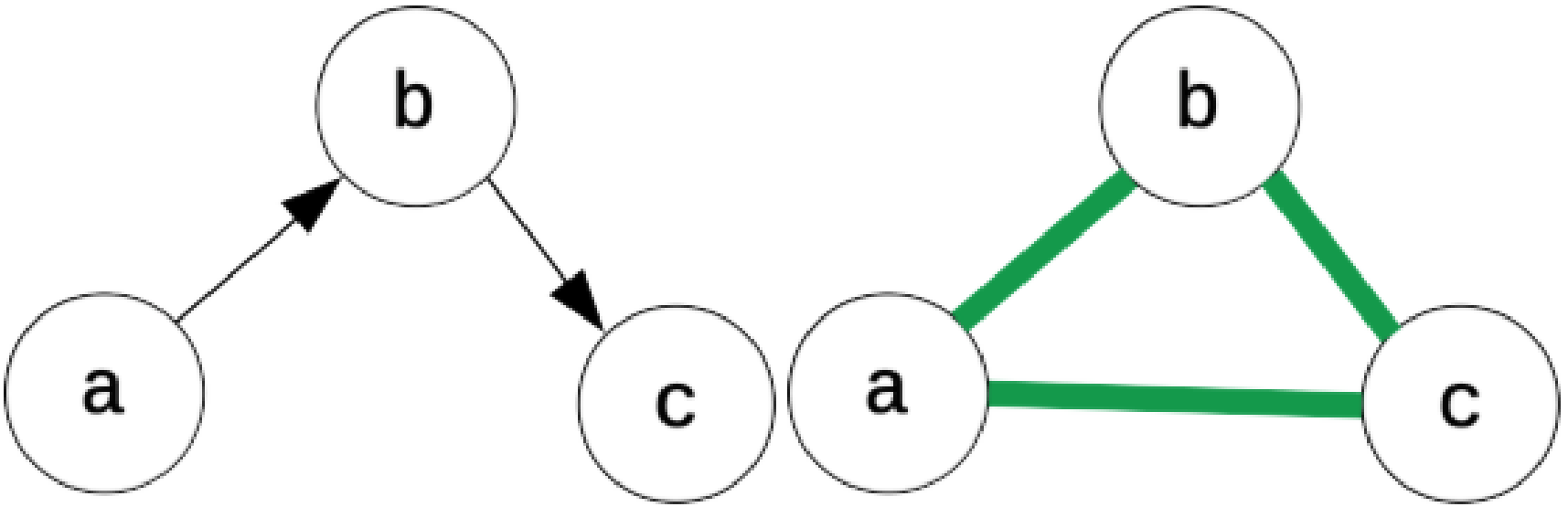
Rule based

...

# formalisms - statistical models




based on correlations. undirectional.



# Differential Equations VS Boolean network models

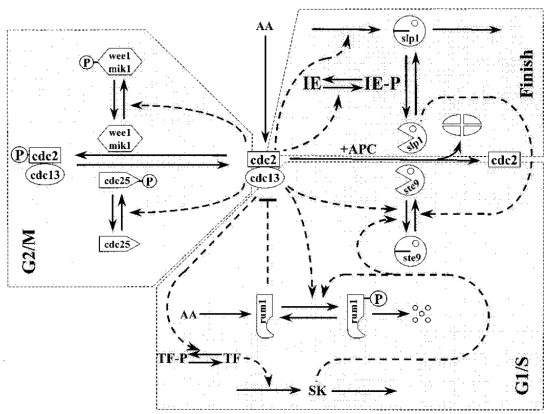



FIG. 2. The wiring diagram of the fission-yeast cell-cycle engine. In the middle of the diagram is Cdc2/Cdc13 (MPF), which is regulated by proteolysis of the Cdc13 component, phosphorylation of Cdc2 subunit, and stoichiometric inhibition of the complex. These processes are arranged according to the cell cycle transitions in which they are involved.

TABLE I. The differential and algebraic equations describing the wiring diagram in Fig. 2.

$$\frac{d[\mathrm{Cdc13}_T]}{dt} = k_1 M - (k_2' + k_2''[\mathrm{Ste9}] + k_2'''[\mathrm{Slp1}])[\mathrm{Cdc13}_T], \quad (1)$$

$$\frac{d[\mathrm{preMPF}]}{dt} = k_{\mathrm{wee}}([\mathrm{Cdc13}_T] - [\mathrm{preMPF}]) - k_{25}[\mathrm{preMPF}] - (k_2' + k_2''[\mathrm{Ste9}] + k_2'''[\mathrm{Slp1}])[\mathrm{preMPF}], \quad (2)$$

$$\frac{d[\mathrm{Ste9}]}{dt} = (k_3' + k_3''[\mathrm{Slp1}])\frac{1-[\mathrm{Ste9}]}{J_3 + 1 - [\mathrm{Ste9}]} - (k_4'[\mathrm{SK}] + k_4[\mathrm{MPF}])\frac{[\mathrm{Ste9}]}{J_4 + [\mathrm{Ste9}]}, \quad (3)$$

$$\frac{d[\mathrm{Slp1}_T]}{dt} = k_5' + k_5''\frac{[\mathrm{MPF}]^4}{J_5^4 + [\mathrm{MPF}]^4} - k_6[\mathrm{Slp1}_T], \quad (4)$$

$$\frac{d[\mathrm{Slp1}]}{dt} = k_7[\mathrm{IEP}]\frac{[\mathrm{Slp1}_T] - [\mathrm{Slp1}]}{J_7 + [\mathrm{Slp1}_T] - [\mathrm{Slp1}]} - k_8\frac{[\mathrm{Slp1}]}{J_8 + [\mathrm{Slp1}]} - k_6[\mathrm{Slp1}], \quad (5)$$

$$\frac{d[\mathrm{IEP}]}{dt} = k_9[\mathrm{MPF}]\frac{1-[\mathrm{IEP}]}{J_9 + 1 - [\mathrm{IEP}]} - k_{10}\frac{[\mathrm{IEP}]}{J_{10} + [\mathrm{IEP}]}, \quad (6)$$

$$\frac{d[\mathrm{Rum1}_T]}{dt} = k_{11} - (k_{12} + k_{12}'[\mathrm{SK}] + k_{12}''[\mathrm{MPF}])[\mathrm{Rum1}_T], \quad (7)$$

$$\frac{d[\mathrm{SK}]}{dt} = k_{13}[\mathrm{TF}] - k_{14}[\mathrm{SK}], \quad (8)$$

$$\frac{dM}{dt} = \mu M, \quad (9)$$

$$[\mathrm{Trimer}] = \frac{2[\mathrm{Cdc13}_T][\mathrm{Rum1}_T]}{\Sigma + \sqrt{\Sigma^2 - 4[\mathrm{Cdc13}_T][\mathrm{Rum1}_T]}}, \quad (10)$$

$$[\mathrm{MPF}] = \frac{([\mathrm{Cdc13}_T] - [\mathrm{preMPF}])([\mathrm{Cdc13}_T] - [\mathrm{Trimer}])}{[\mathrm{Cdc13}_T]}, \quad (11)$$

$$[\mathrm{TF}] = G(k_{15}M, k_{16}' + k_{16}''[\mathrm{MPF}], J_{15}, J_{16}), \quad (12)$$

where

$$k_{\mathrm{wee}} = k_{\mathrm{wee}}' + (k_{\mathrm{wee}}'' - k_{\mathrm{wee}}')G(V_{\mathrm{awee}}, V_{\mathrm{iwee}}[\mathrm{MPF}], J_{\mathrm{awee}}, J_{\mathrm{iwee}}),$$

$$k_{25} = k_{25}' + (k_{25}'' - k_{25}')G(V_{a25}[\mathrm{MPF}], V_{i25}, J_{a25}, J_{i25}),$$

$$\Sigma = [\mathrm{Cdc13}_T] + [\mathrm{Rum1}_T] + K_{\mathrm{diss}},$$

$$G(a,b,c,d) = \frac{2ad}{b - a + bc + ad + \sqrt{(b - a + bc + ad)^2 - 4ad(b-a)}}.$$

TABLE II. Parameter values for wild-type cells. All constants have units $\mathrm{min}^{-1}$, except the $J$s, which are dimensionless Michaelis constants, and $K_{\mathrm{diss}}$, which is a dimensionless equilibrium constant for trimer dissociation.

Cdc13 synthesis and degradation:
$k_1 = 0.03$, $k_2' = 0.03$, $k_2'' = 1$, $k_2''' = 0.1$.
Ste9 activation and inactivation:
$k_3' = 1$, $k_3'' = 10$, $J_3 = 0.01$, $k_4' = 2$, $k_4 = 35$, $J_4 = 0.01$.
Slp1 synthesis, degradation, activation and inactivation:
$k_5' = 0.005$, $k_5'' = 0.3$, $k_6 = 0.1$, $J_5 = 0.3$
$k_7 = 1$, $k_8 = 0.25$, $J_7 = 0.001$, $J_8 = 0.001$.
IE activation and inactivation:
$k_9 = 0.1$, $k_{10} = 0.04$, $J_9 = 0.01$, $J_{10} = 0.01$.
Rum1 synthesis, degradation and inhibition:
$k_{11} = 0.1$, $k_{12} = 0.01$, $k_{12}' = 1$, $k_{12}'' = 3$, $K_{\mathrm{diss}} = 0.001$.
SK synthesis and degradation:
$k_{13} = 0.1$, $k_{14} = 0.1$.
TF activation and inactivation:
$k_{15} = 1.5$, $k_{16}' = 1$, $k_{16}'' = 2$, $J_{15} = 0.01$, $J_{16} = 0.01$.
Wee1 activation and inactivation:
$V_{\mathrm{awee}} = 0.25$, $V_{\mathrm{iwee}} = 1$, $J_{\mathrm{awee}} = 0.01$, $J_{\mathrm{iwee}} = 0.01$.
Cdc25 activation and inactivation:
$V_{a25} = 1$, $V_{i25} = 0.25$, $J_{a25} = 0.01$, $J_{i25} = 0.01$.
Rate of tyr-phosphorylation and dephosphorylation:
$k_{\mathrm{wee}}' = 0.15$, $k_{\mathrm{wee}}'' = 1.3$, $k_{25}' = 0.05$, $k_{25}'' = 5$.
Growth rate:
$\mu = 0.005$.

*'Mathematical model of the cell division cycle of fission yeast'. Chaos: An Interdisciplinary Journal of Nonlinear Science. Novak et al. 2001*

# Differential Equations VS Boolean network models

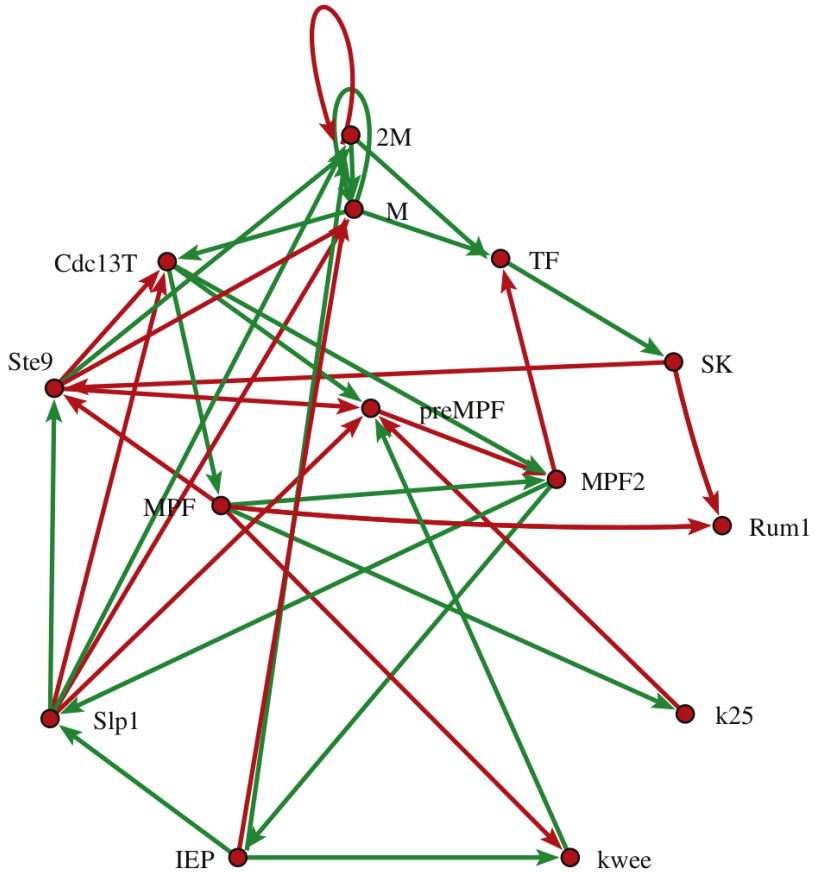
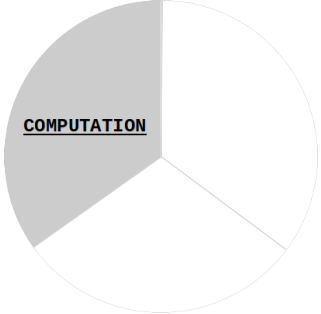



**Fig. 2.** Interaction network of the Boolean model. Green links correspond to $T_{ik} = +1$ and red links to $T_{ik} = -1$.

*« The transition from differential equations to Boolean networks: A case study in simplifying a regulatory network model ». M. Davidich & S. Bornholdt. Journal of Theoretical Biology. Dec 2008.*

# Eg of tool : COPASI

Complex Pathway Simulator




runnabled models from a diagram transformed in ordinary differential equation.

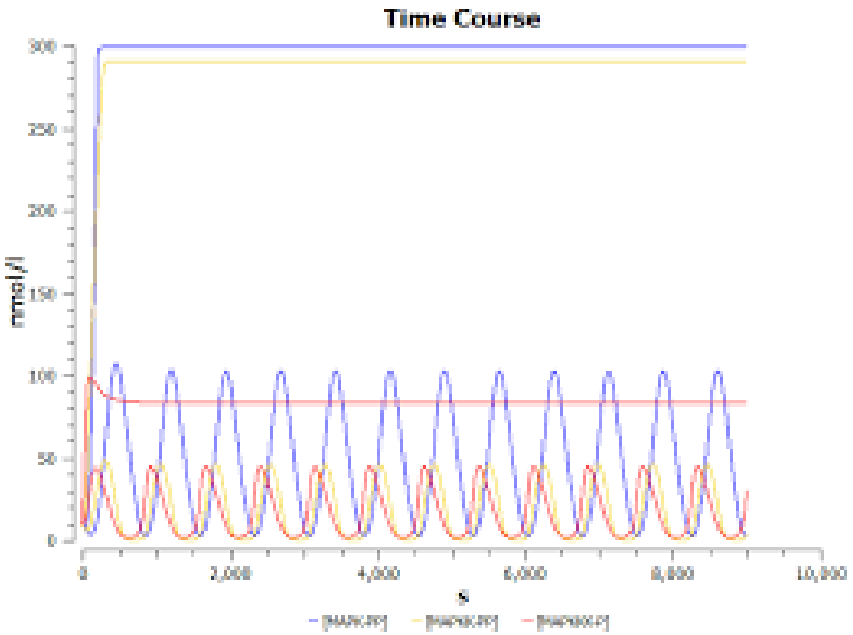



Figure 2: from “Building and Simulating Models using COPASI”, © 2016, Nicolas LeNovère, Viji Chelliah, Bhupinder Virk. creative commons Attribution - Share Alike 4.0 licence.

# Model analysis. Eg. attractors

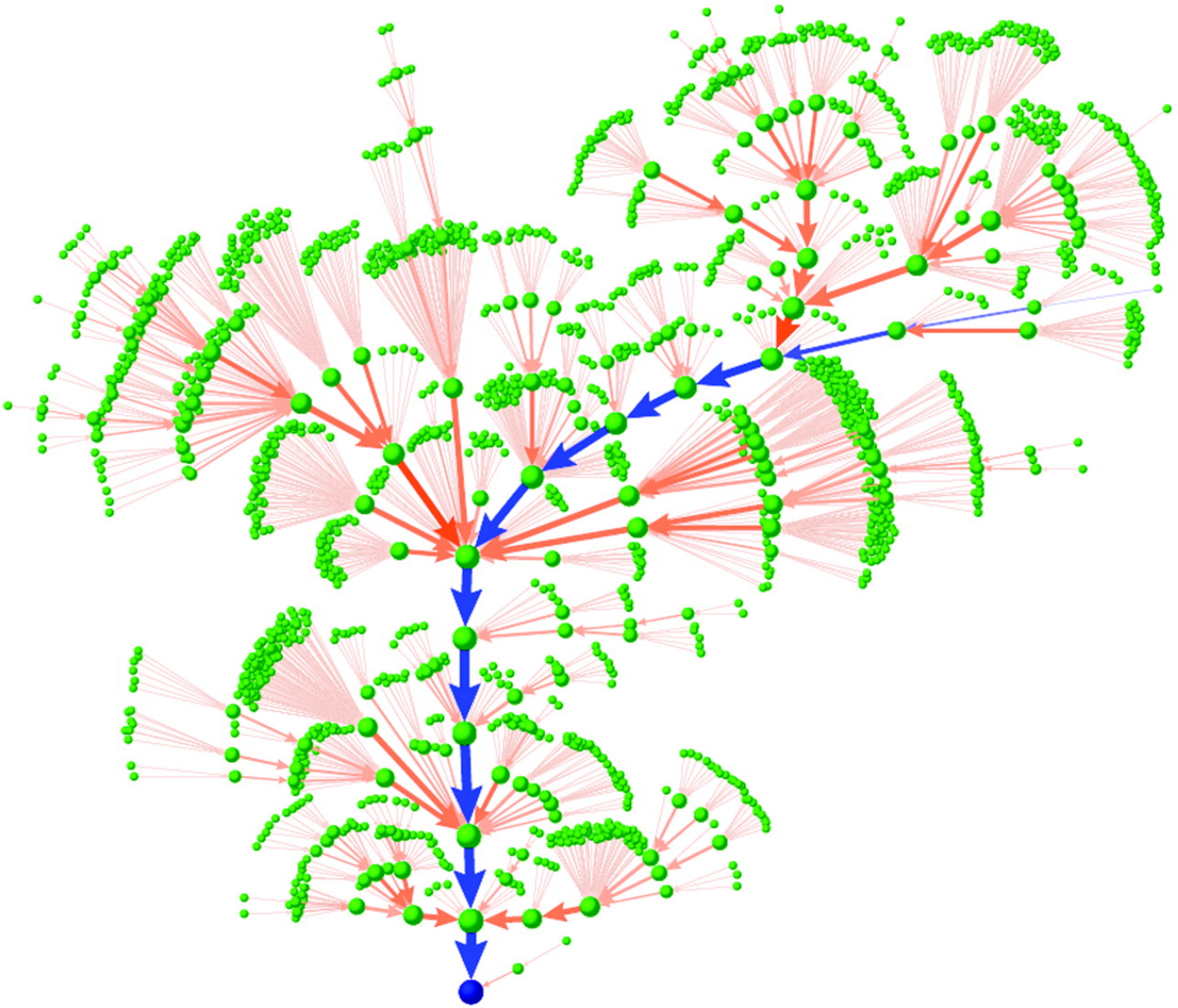
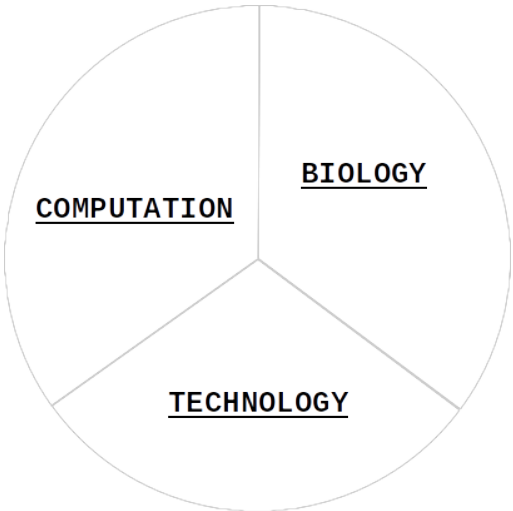


Figure 3: A model that fits biological reality.

# systems biology triade cycle

- long and tedious
- lot of challenges at each steps
- standards needed to allow model sharing (and tools to communicate)




the combine
computational modeling in biology network

SBML™

BioModels

# What is available in BioModels ?

- models from litterature.

| | non-curated (7894) | curated (718) |
|---|---|---|
| ODE | 583 | 190 |
| Constraint-based | 129 | |
| Logical | 14 | 3 |
| Petri net | 3 | 1 |

From Build: e6ea664 | *Mon, 15 Oct 2018 14:10:54 +0100**

- path2models :

models automatically derived from pathway resources, such as `KEGG`, `BioCarta`, `MetaCyc`, `PID` and `SABIO-RK`.

- Metabolic (112,898 models)
- Non-metabolic (27,531 models)
- Whole genome metabolism (2,641 models)

# And... there is BioModels inside LOD ! :D

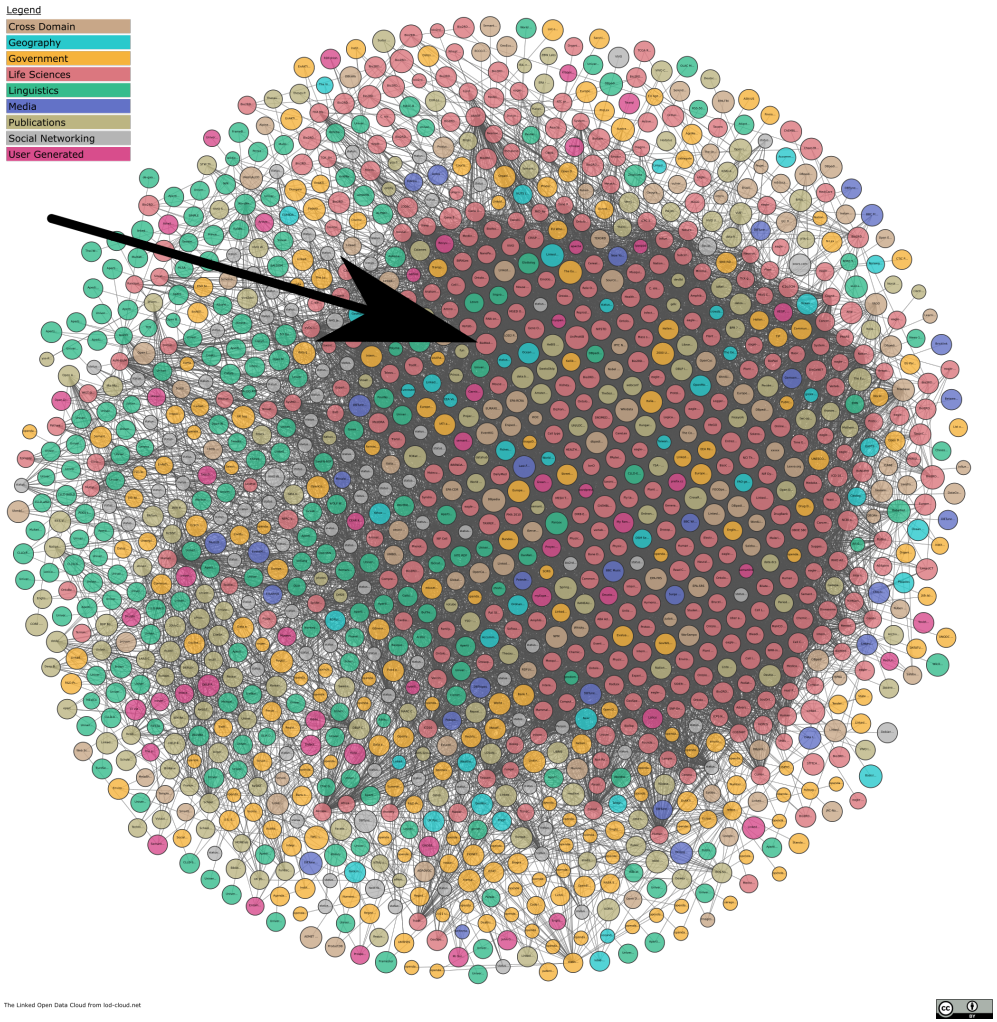
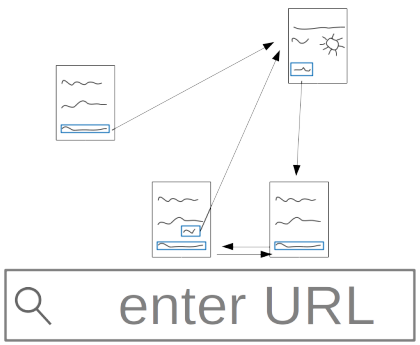
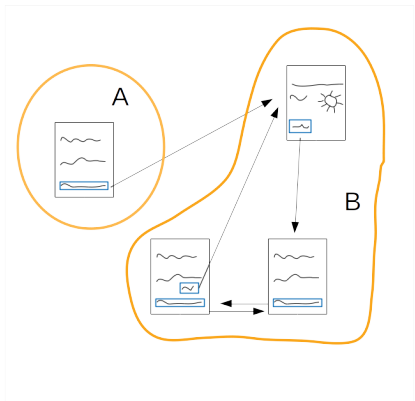
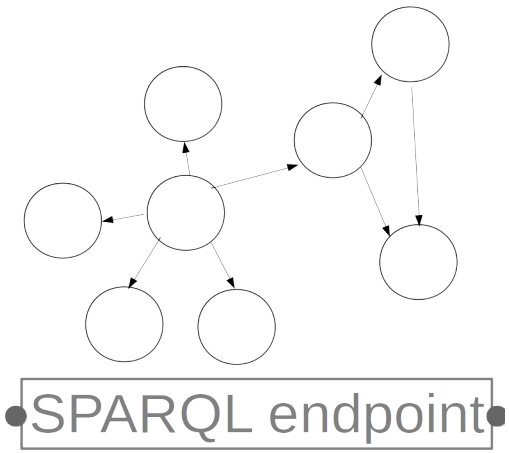



Figure 4: 1,229 datasets with 16,125 links *(June 2018)*

# a quick history of the web . . .




from **hyperlinks** to **applications** to **semantic web**

# Web 3.0 : W3C Semantic Web

*The Semantic Web will bring structure to the meaningful content of Web pages. (Tim Bernard Lee et al. Scientific American - Feature Article: The Semantic Web. May 2001)*

- **data** (not files) are **connected** :
  L[O]D : Linked [Open] Data (graph database) ;
  `RDF (objet, predicat, sujet)`
- data is given **well-defined meaning** :
  semantics, ontologies `RDFS / OWL`

. . . in order for **computer** programs
to **manipulate** data
meaningfully and **automatically**.

# if everyone plays the game... data interoperablity <3

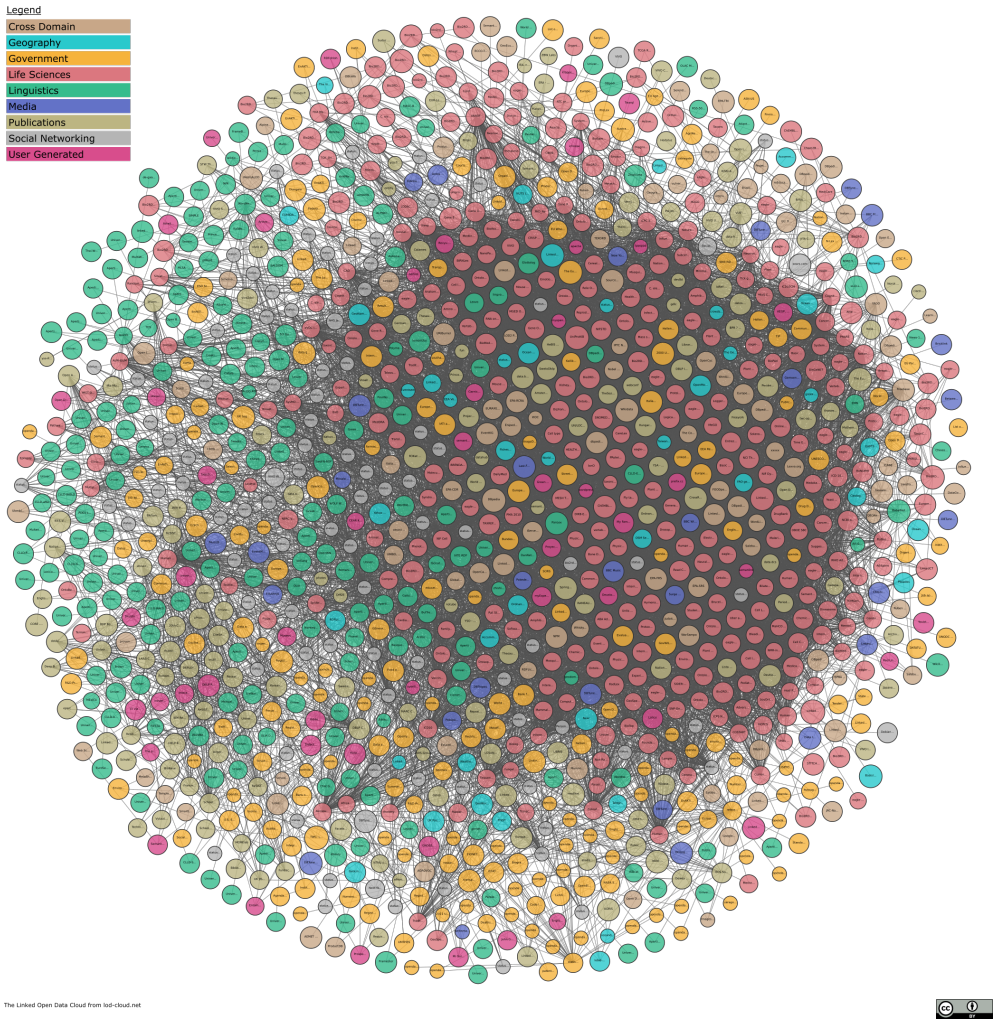
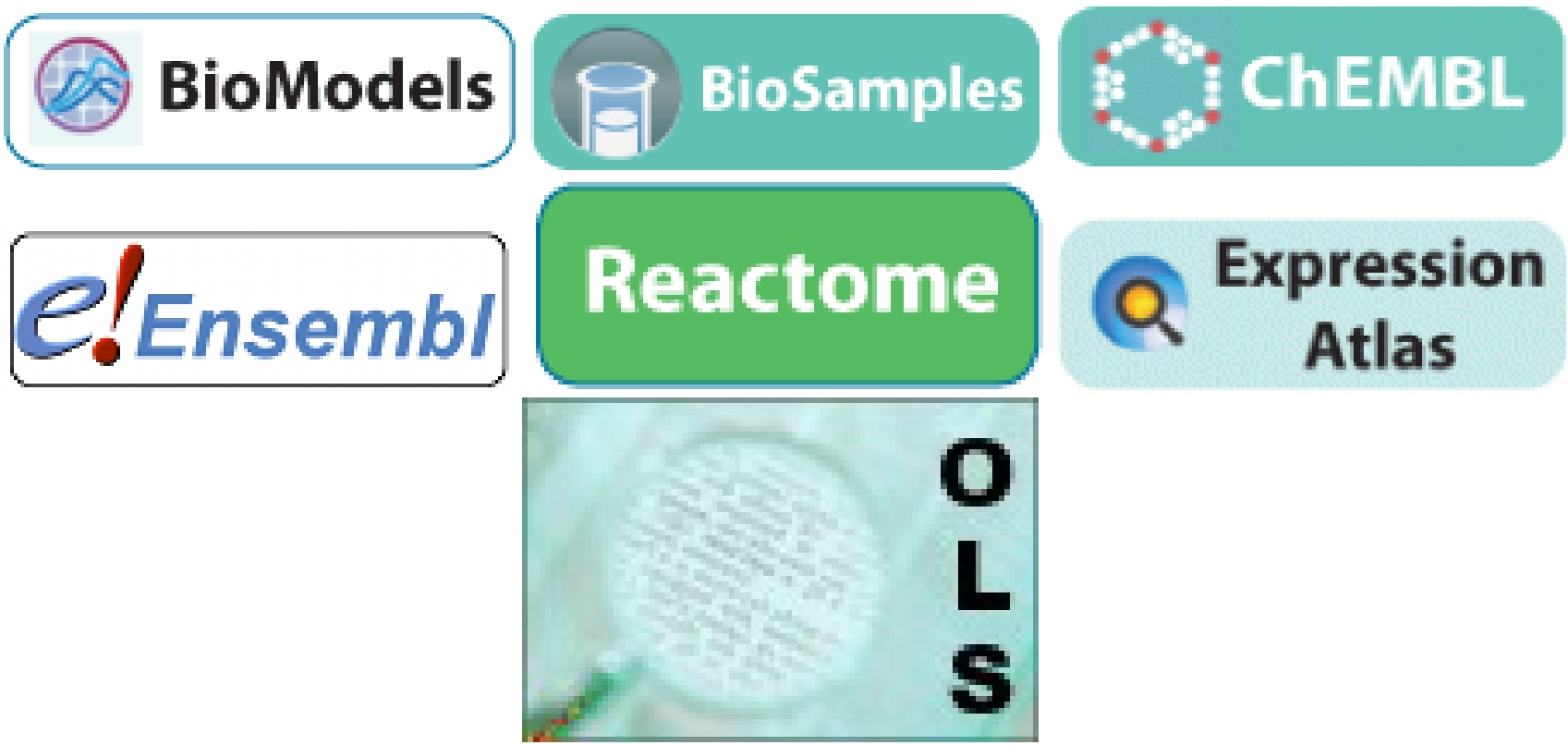


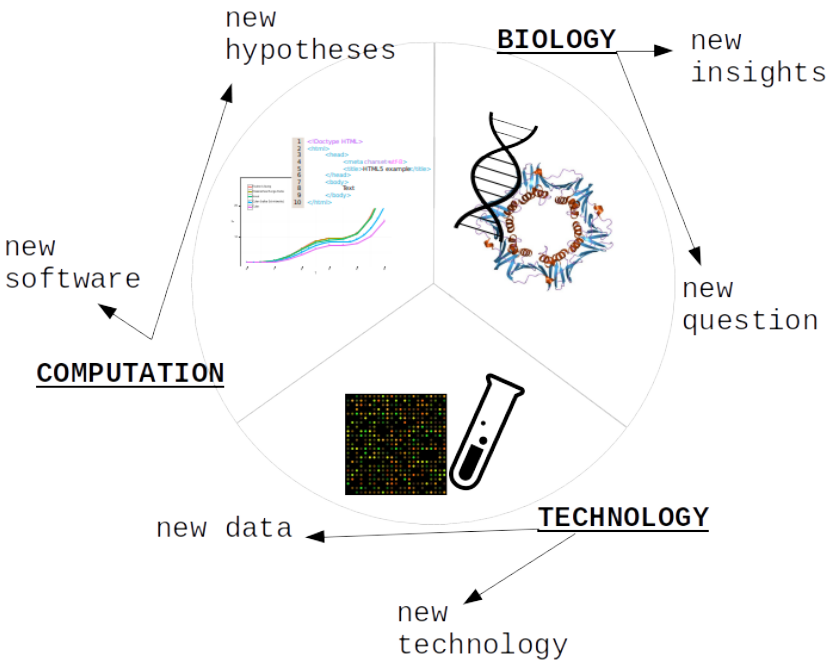

Figure 5: 1,229 datasets with 16,125 links *(June 2018)*

# at total : 7 EBI datasets in the LOD




(OntologyLookupService)

# in summary :




+

W3C Semantic Web

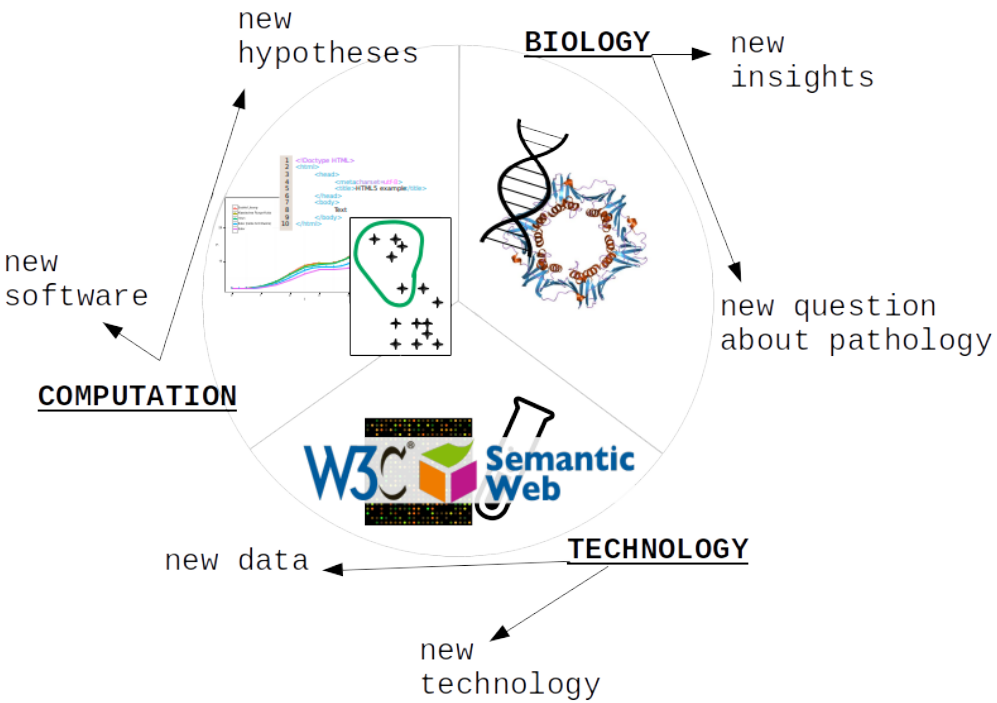
BioModels BioSamples ChEMBL

e!Ensembl Reactome BioSamples

OLS

# Goal: upgrade the SB models developpement cycle

some interesting points to be explored :




- biomodical context = pathology
- use existing data to build possible models
- confront the models with the biomedical knowledge
- simulate models
- select the best one(s)

# Currently :

We investigate some work dealing with biomodels LOD dataset.

Ron Henkel, Robert Hoehndorf, Tim Kacprowski, Christian Knüpfer, Wolfram Liebermeister, Dagmar Waltemath; Notions of similarity for systems biology models, Briefings in Bioinformatics, Volume 19, Issue 1, 1 January 2018, Pages 77–88, https://doi.org/10.1093/bib/bbw090

Thanks for your attention. . .
Any questions ? :)

# References I

# Annexes

# Comparison of formalisms commonly used for a specific purposes

# Comparison of PBN and DBN for inferring GRN from time series gene expression data

*« Comparison of Probabilistic Boolean Network (PBN) and Dynamic Bayesian Network (DBN) approaches for inferring gene regulatory networks (GRN) » Li et al. BMC Bioinformatics. nov 2007.*

- both have good performances in modeling the GRNs
- the accuracy in terms of recall and precision can be improved if a smaller subset of genes is selected for inferring GRNs. (if more genes are selected for inferring GRNs, the network contains more edges and it is more difficult to successfully identity the interactions among genes)
- the accuracy of both approaches is dependent upon the number of selected genes and time points of gene samples.
- DBN identify more gene interaction than PBN
- DBN is more time consuming