

MBI MASTER'S DEGREE
PROGRAMME IN BIOINFORMATICS

Medicel
RESEARCH REDEFINED

Christophe Roos – PhD, docent, Principal Analyst, christophe.roos@medicel.com

Company presentation: Medicel

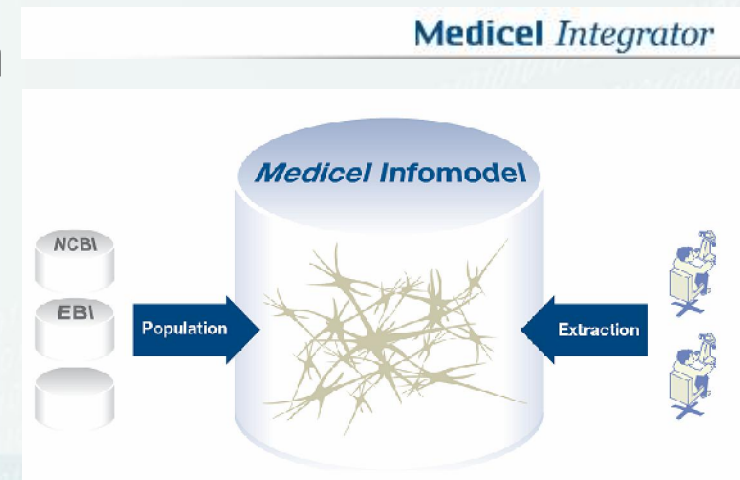
IT platform for integrative research

Medicel Oy – Quick Facts

- Privately held, estd. 2001
- Headquartered in Espoo, Finland
- ~60 person team worldwide
 - 30 software engineers
 - 20 other engineers, mathematicians, statisticians, *etc.*
 - 5 lab technicians
 - 2 PhD, 1 MD
- Entering market with own products (IP, filed patents)
- Out-licensing IT technology
- Selling high-throughput instruments

Medicel Oy – Technology

- Supplier for research programs in the fields of medicine, pharmacology, biology, agronomy, ...
- 3 product lines
 - Integrator software
 - Data
 - Methodology
 - Project
 - High throughput multi-well bioreactor
 - Services



Medicel Oy – Aims

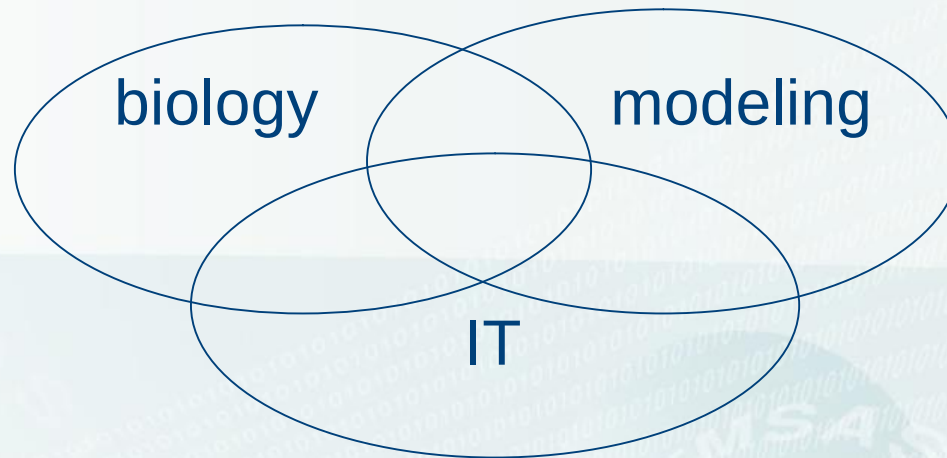
Medicel Integrator

- Maximise computer readability – accelerate research
- Formalise knowledge representation – data model
- Centralise data – integrate data silos
- Broaden understanding – multi-domain analyses
- Support discovery process – modeling & simulation
- Get knowledge early – pleiotropic & side effects
- Long-term data accumulation – projects span years

Part 2.a – Technology

To facilitate inter-disciplinary research

Inter - disciplinary



Biological complexity beyond intuition –
use models to produce hypotheses
Wet lab experimentation expensive –
computer simulations cheap

Research layer

Project management platform

Integration layer

Biological applications for data refinement

NMR, X-ray

Imaging

Mass spec

DNA sequencing

RNA quantisation

Genotyping

Chromatin IP

Molecular biology

Literature

Public databases

Measurement layer

Core technology platforms

I T layer

Data warehouse

Data management platform



→ Data -> Model -> Predict -> Measure



Experimental

Theoretical

Data



Experimental

Modeling

Data Integration

It's about communication



Experimental

Theoretical

Data

Very different epistemological cultures in *physics* and *biology*



Experimental

Modeling

Data Integration

Part 2.a – Technology

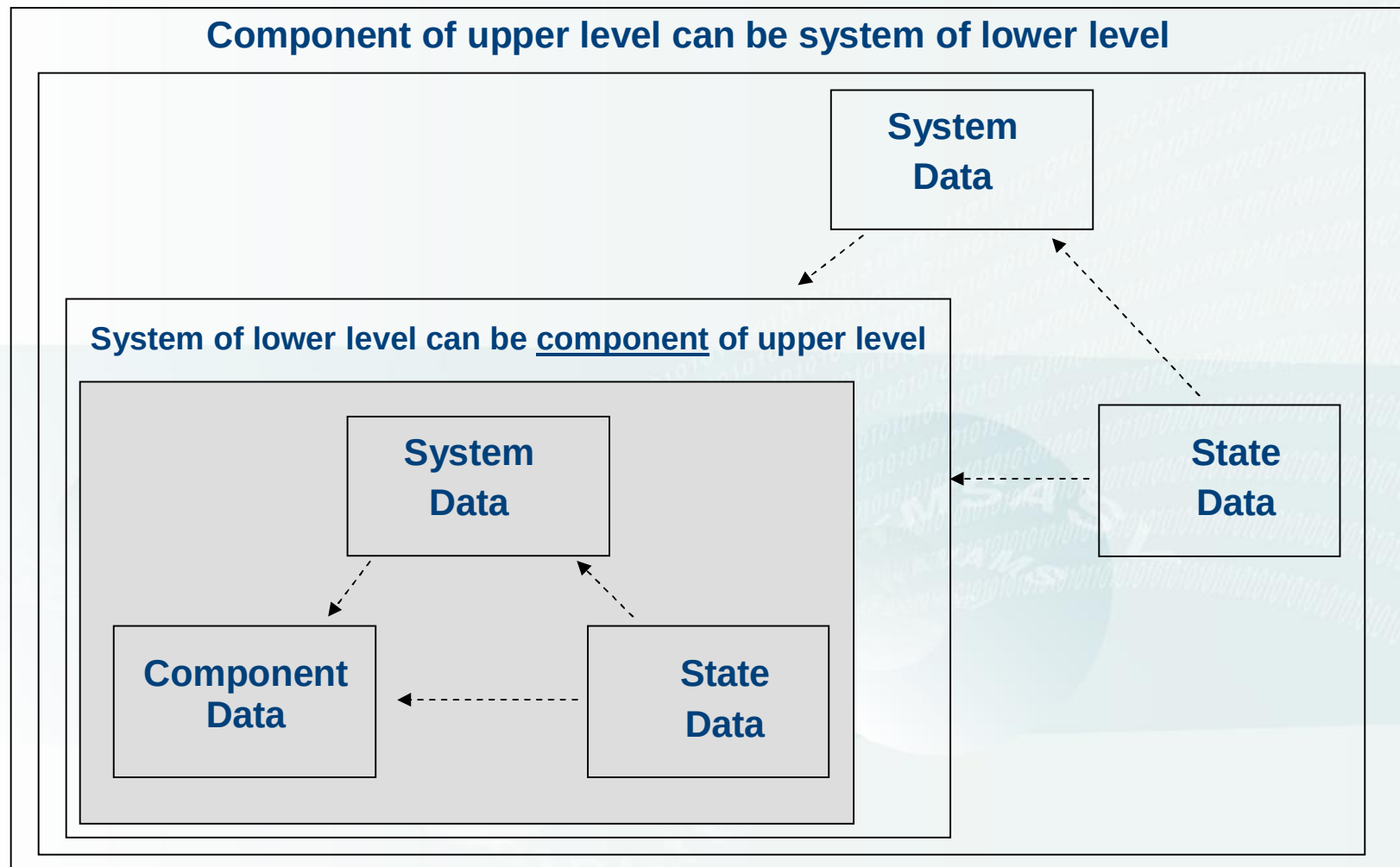
To teach and communicate biological system descriptions

To describe biological systems

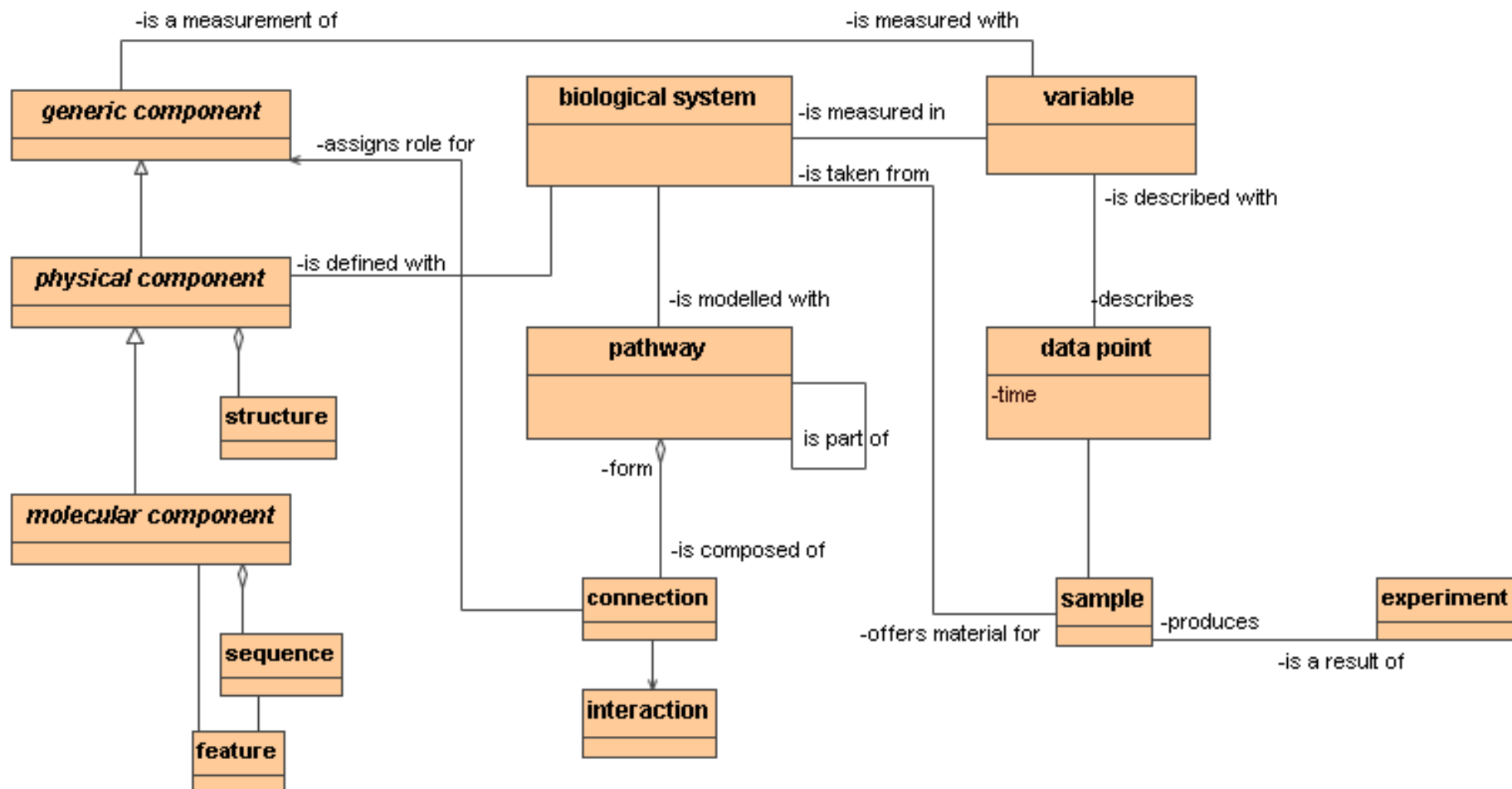
Medicel Integrator

- | Document and convey biological knowledge and understanding in computer readable format
- Use an adequate formalism
 - Components
 - System
 - State
- General solutions help to survive through the years

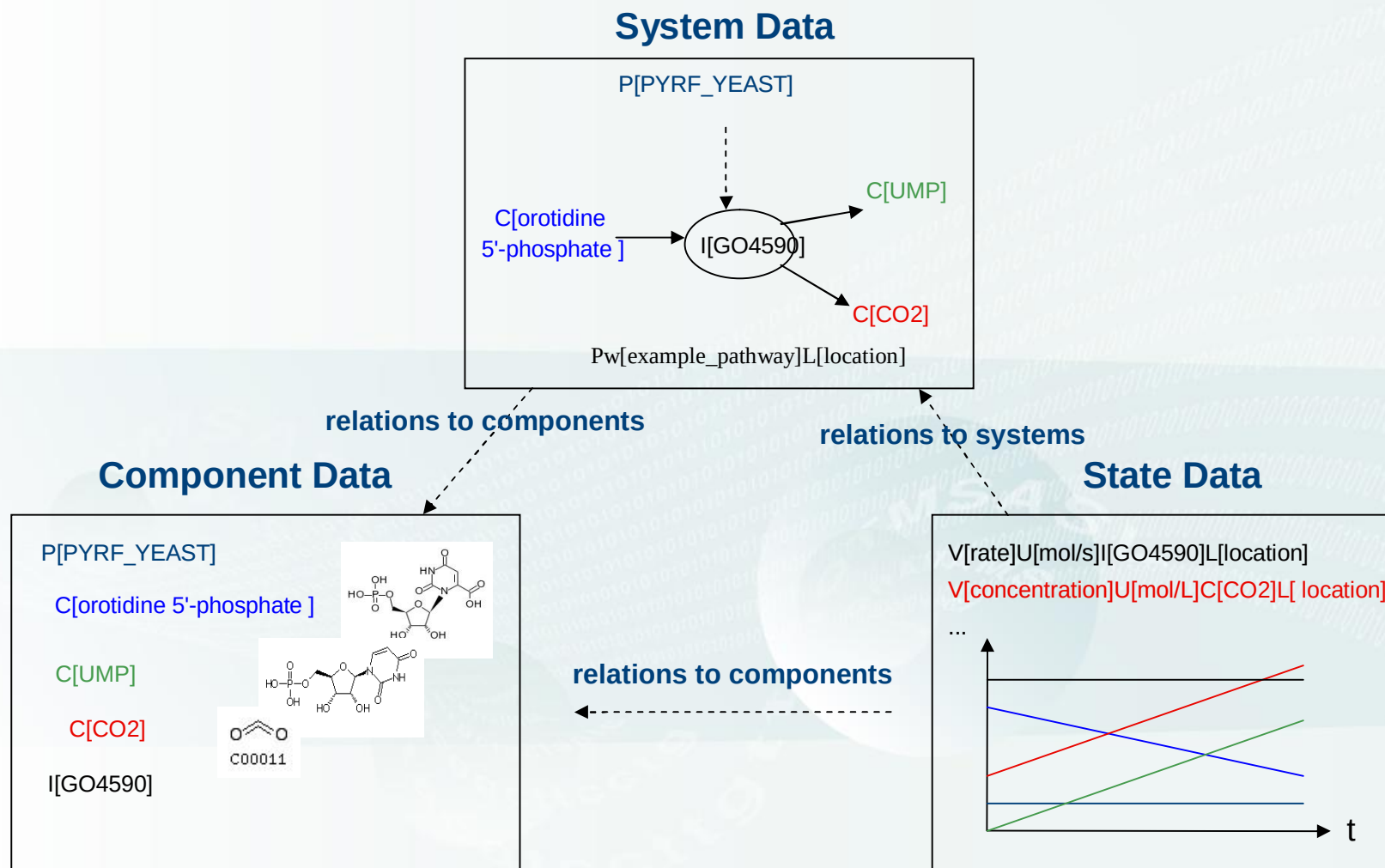
General systems theory



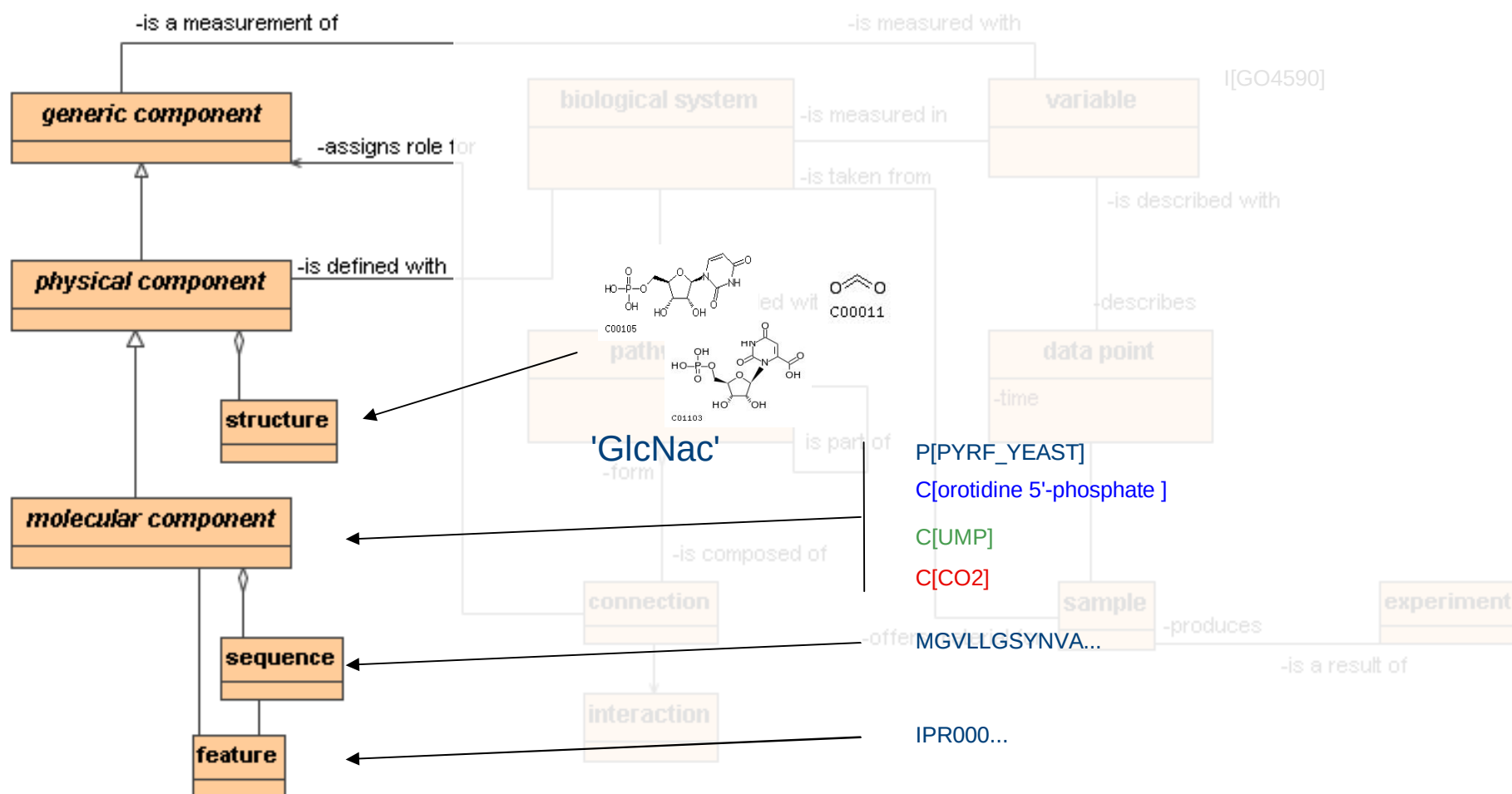
Information model: simple overview



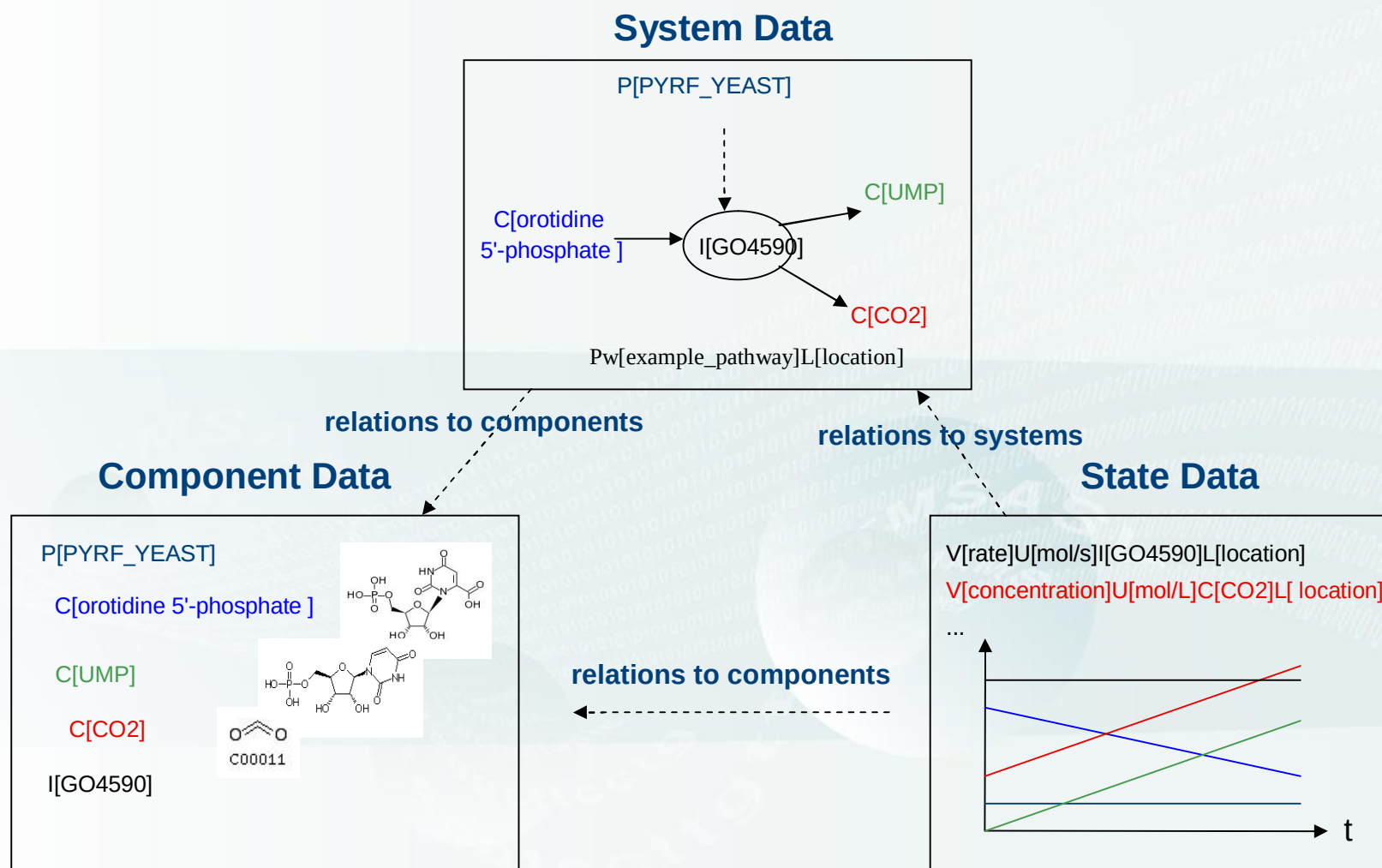
Biological systems



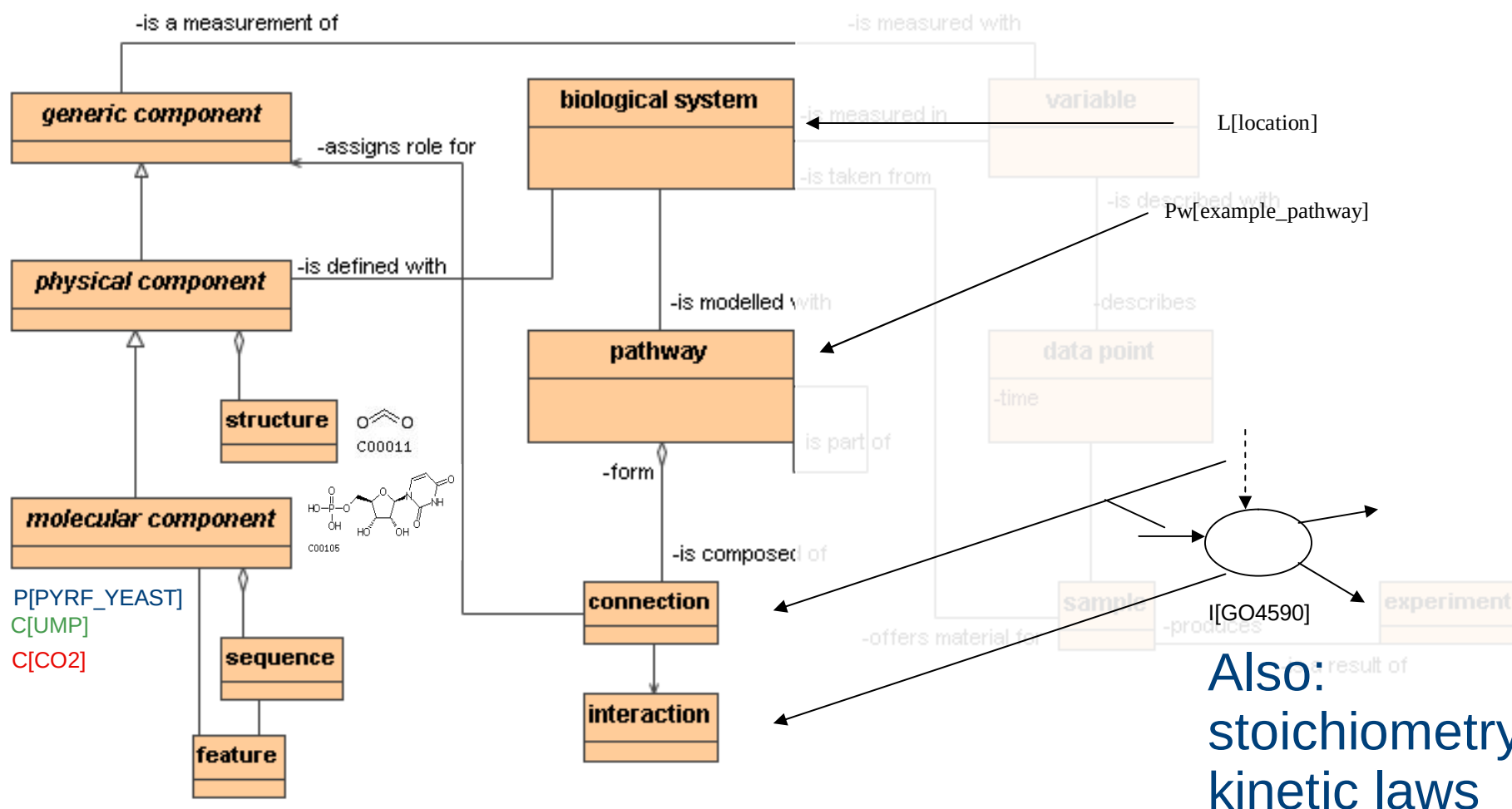
The model for *component*



Biological systems



The model for the *topology*



Sa[gis-035/b]V[concentration]U[mol/L]P[PYRF_YEAST]L[location2]T[t]

V[concentration]U[mol/L]C[UMP]L[location2]T[t]

V[concentration]U[mol/L]C[CO2]L[location2]T[t]

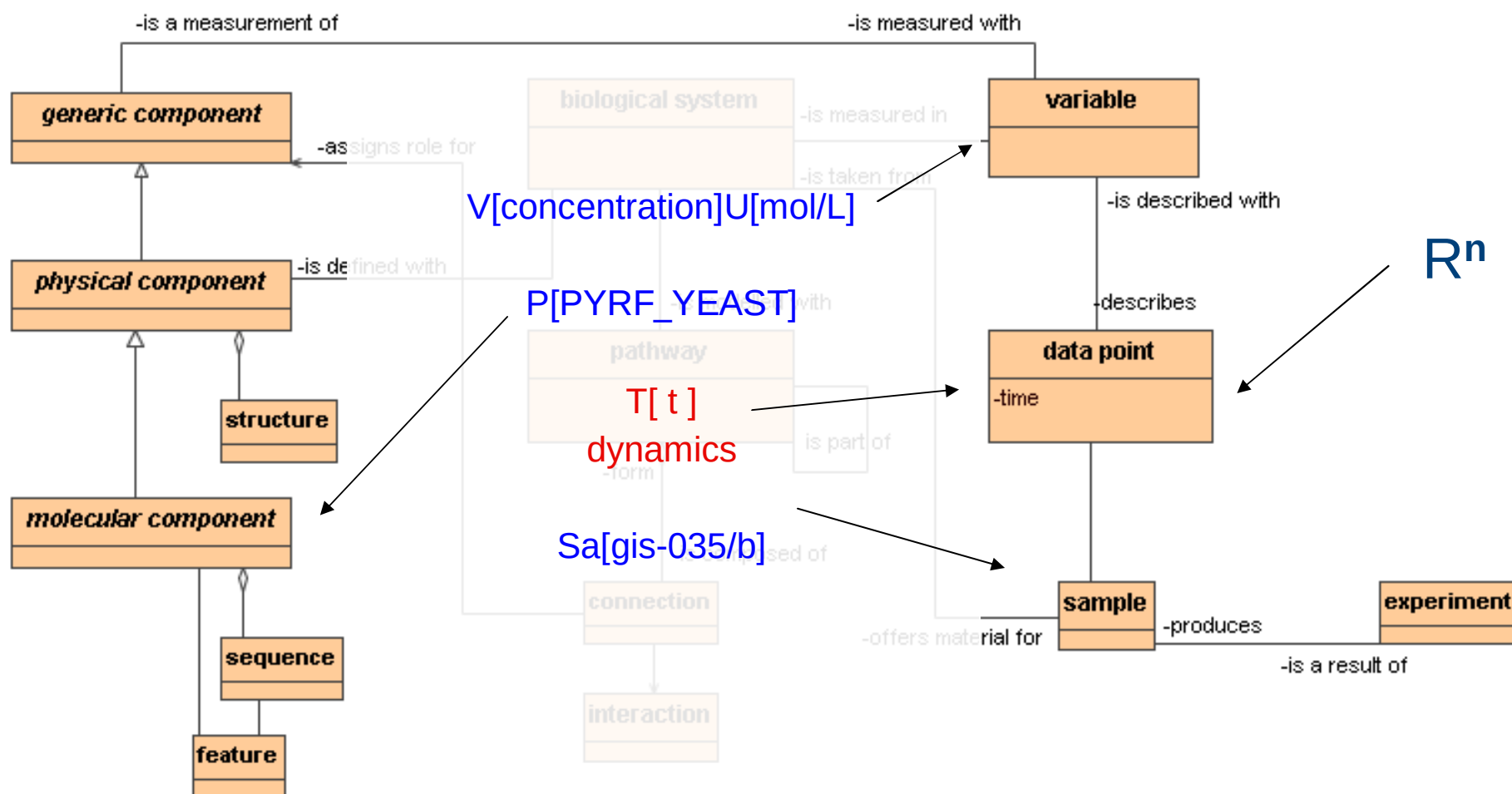
 $\in \mathbb{R}^n$

V[rate]U[mol/s]I[GO4590]L[location2]T[t]

V[volume]U[L]L[location2]T[t]



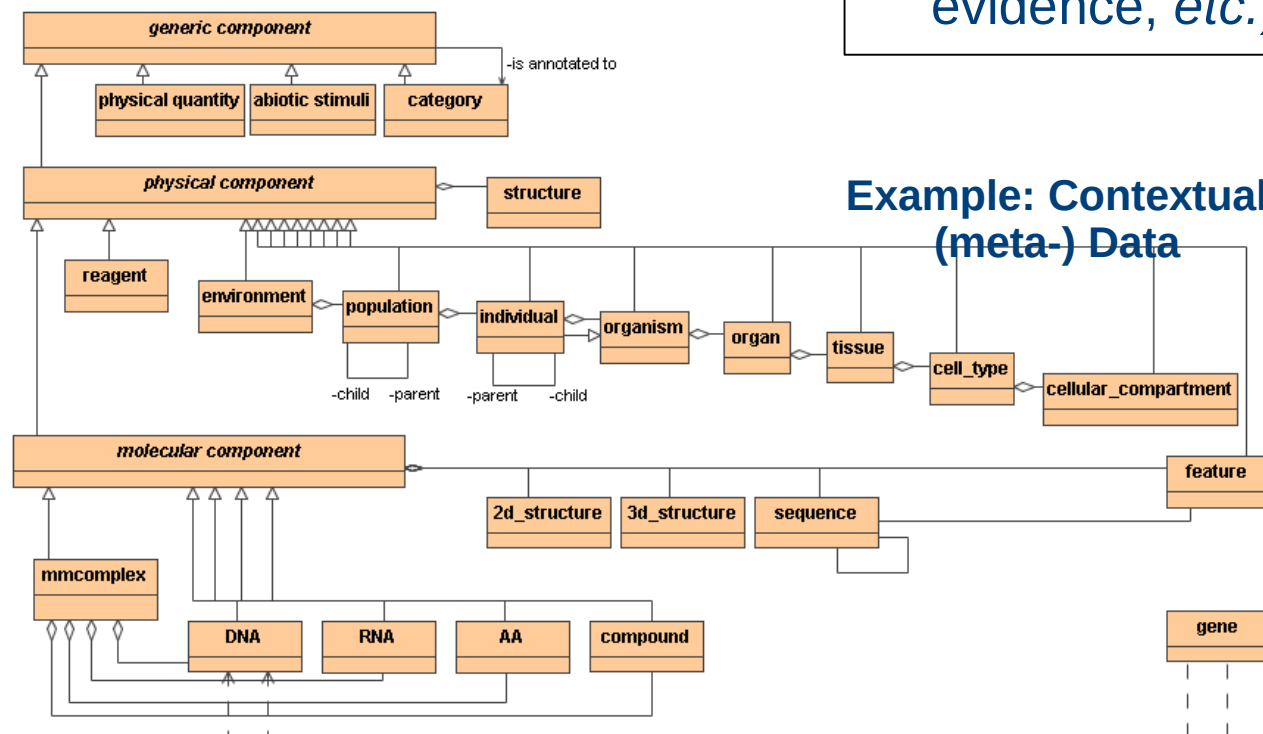
The model for *state*



Information model – generic & flexible

Medicel Integrator

Ca 50 tables covering biological data
Ca 150 tables covering other aspects
(project, owner, version, source,
evidence, etc.)



Part 2.b – Technology

Data integration

Data integration – Data modeling

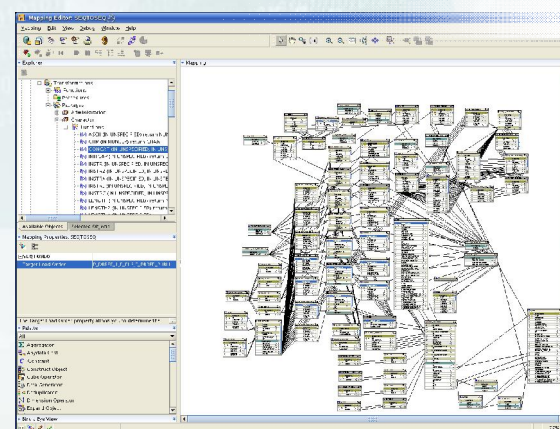
Medicel Integrator

- The information model in a pivotal role:
 - Build a data repository 'at home' and keep it at hand for further processing and knowledge building
 - Import from the most relevant data sources
 - Include measurement data
 - Export structured data to collaborators or to public repositories

Data integration – The IT essentials

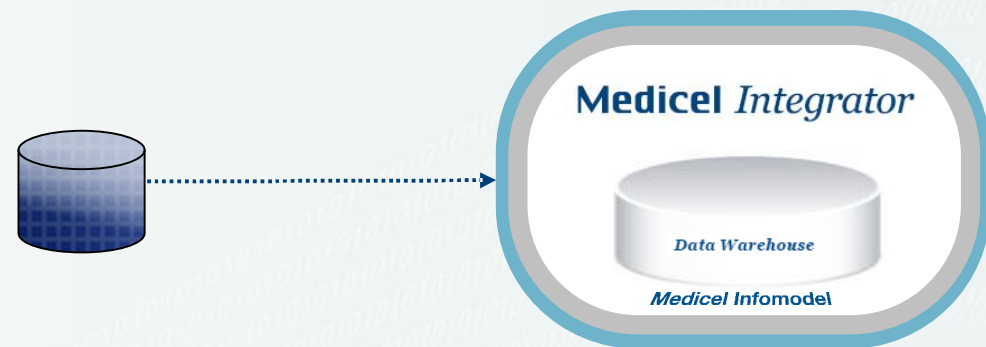
- Schema mapping
- ETL: Extract, transform, load
- Resolve differences (subjective, objective)
- Provenance, evidence codes, ontologies
- Update and maintain multiple versions
- Same for public data and private data

UniProt schema mapping
to Medicel schema



Medicel Integrator

Data integration protocols tested



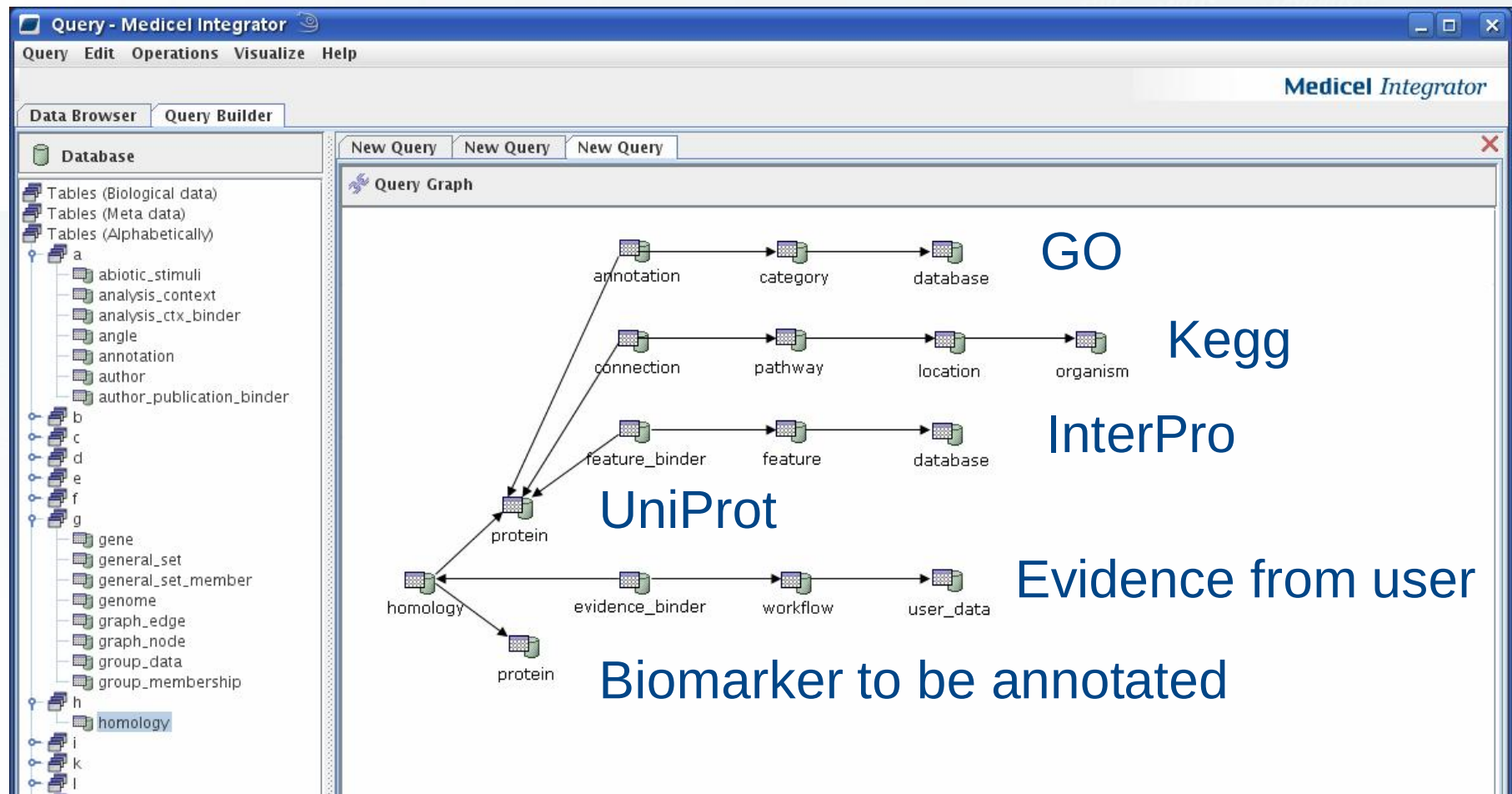
- . ENSEMBL
- . NCBI Taxonomy
- . NCBI Refseq Proteins
- . UniProt/Swissprot
- . UniProt/TrEMBL
- . Interpro
- . Mammalian Phenotype Ontology
- . KEGG*
- . Human Disease Ontology

- . IntAct
- . GO
- . Cell Ontology
- . Chebi
- . Cytomer
- . Brenda Tissue Ontology
- . PDB
- . PubMed

- . ICD9
- . NCBI Refseq Genomes
- . NCBI RefSeq RNA
- . OMIM
- . SNOMED
- . EMBL
- . HapMap
- . GEO
- . ArrayExpress
- . Etc.

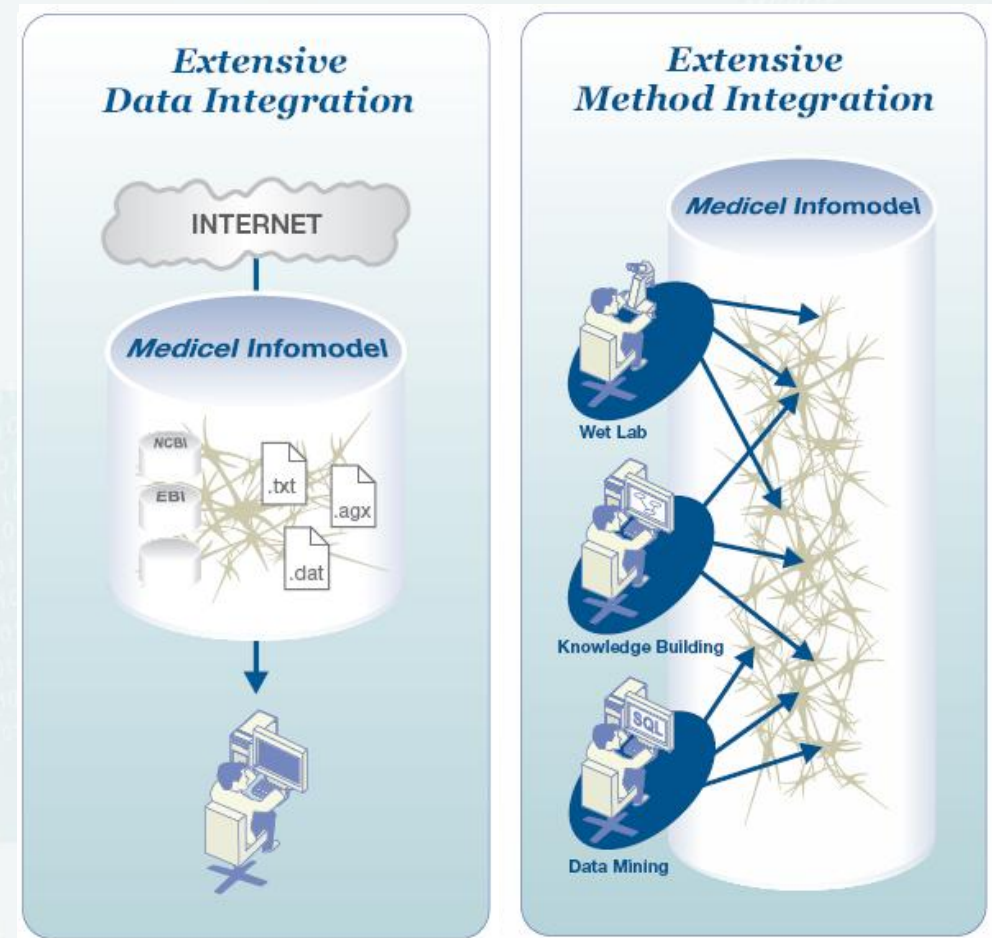
Data integration – Complex queries

- Comprehensive searches: annotate using homology



Integrated data -> tools -> processes

- Standards, formats
- Scientific process
 - Biological context
 - Applied methodologies
- Accumulate
 - Data
 - Methods
- Communicate
 - Conclusions
 - Hypotheses



Part 2.c – Technology

Projects and data modules (versioning)

Collaboration platform

Project A

Project C

Gene Ontology

Project B

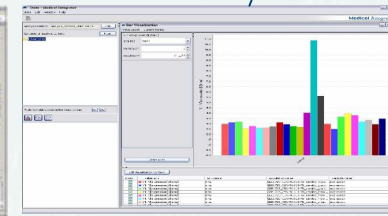
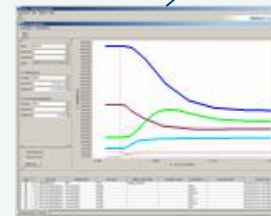
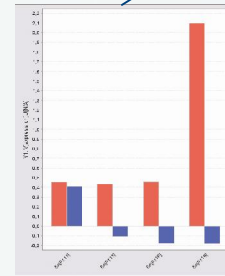
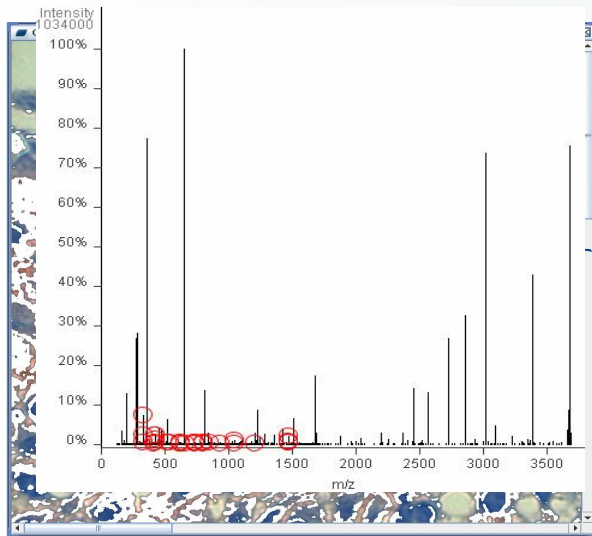
Measurement data

Uniprot

Part 2.d – Technology

Multi-domain analysis

Multidomain analysis



Research project

Pathway Analysis

Microscopy

Generic

Customer Library

Proteomics

Transcriptomics

Measurement technologies

Mass spec

Microscopy

DNA sequencing

Genotyping

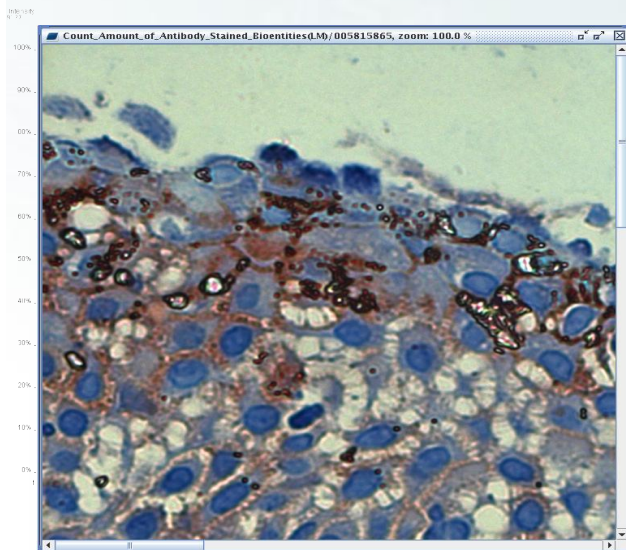
NMR, X-ray

RNA quantisation

Molecular biology

Chormatin IP

30



Part 3 – References

References

- Medicel provides the IT infrastructure in EU FP7 framework projects
 - *HEALTH* projects: large systems biology projects
 - Apoptosis, T-cell differentiation, stem cells
 - *KBBE* projects: Food, agriculture and fisheries
 - Plant parthenogenesis
 - *ICT* projects: Information and Communication Technologies
 - e-Infrastructures, sustainable and personalized healthcare
 - *Leonardo: Training in systems biology*
- Evaluation of infrastructure
 - Academic projects (Australia, Singapore, USA)
 - Pharmaceutical companies

Summary – Research infrastructure

Inter – disciplinary !

Medicel Integrator

