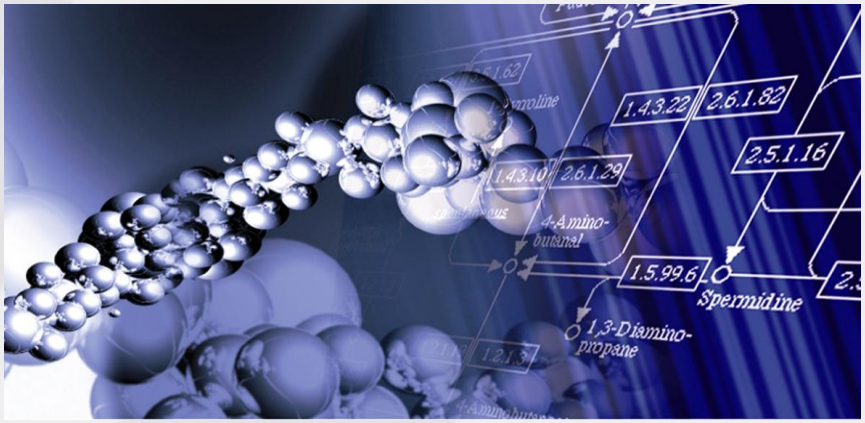
# Data modeling:
# the key to biological data integration




François Rechenmann

NETTAB 2012

Genostar
Bioinformatics solutions

# Biological data: not so big, but highly heterogeneous and evolving

## Big data

- Satellite images, particle physics,…
- Banks, insurance, telecom companies,…

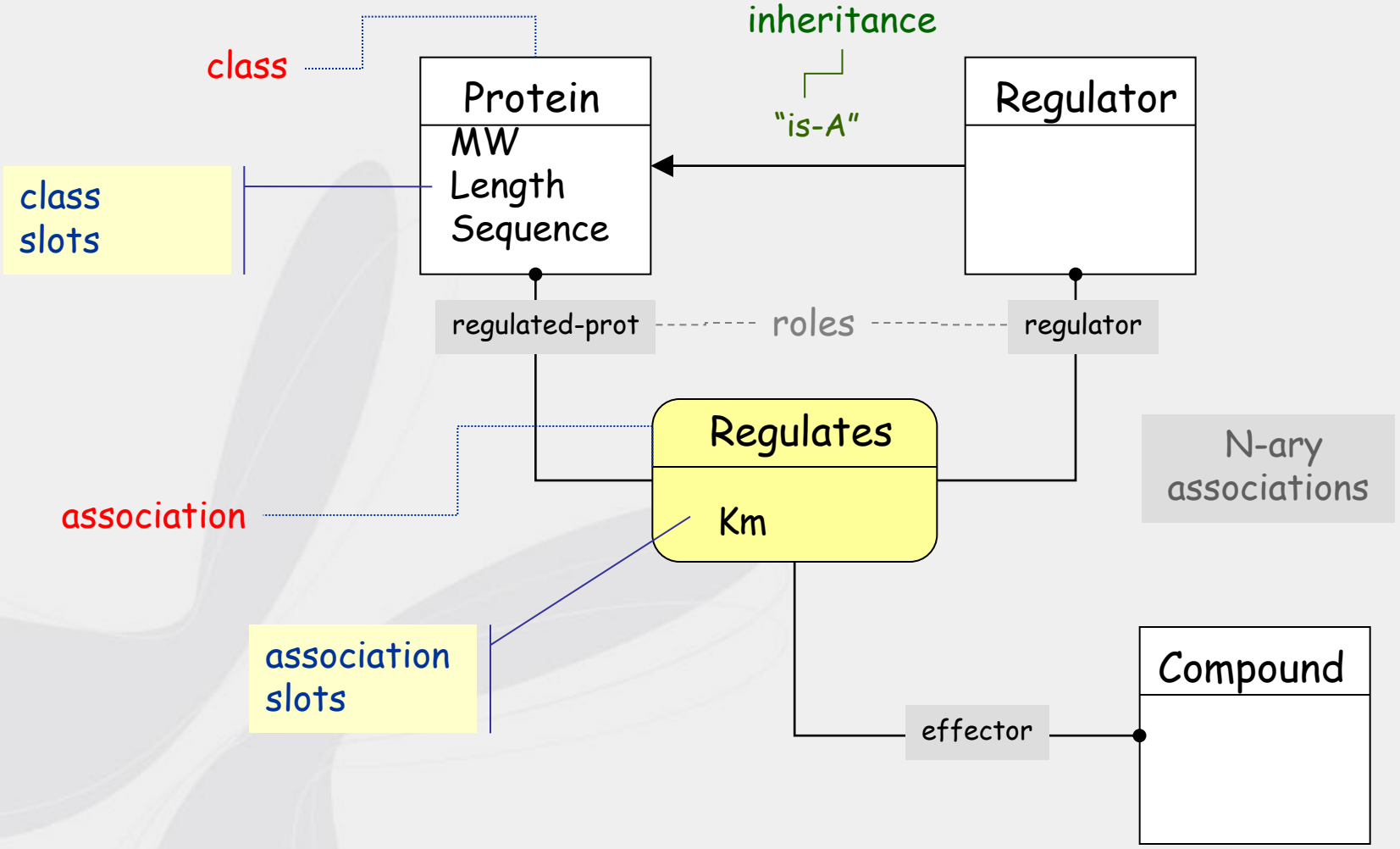
## Heterogeneous biological data

- Genomic, transcriptomic, proteic, metabolic data
- Spectra, structures…

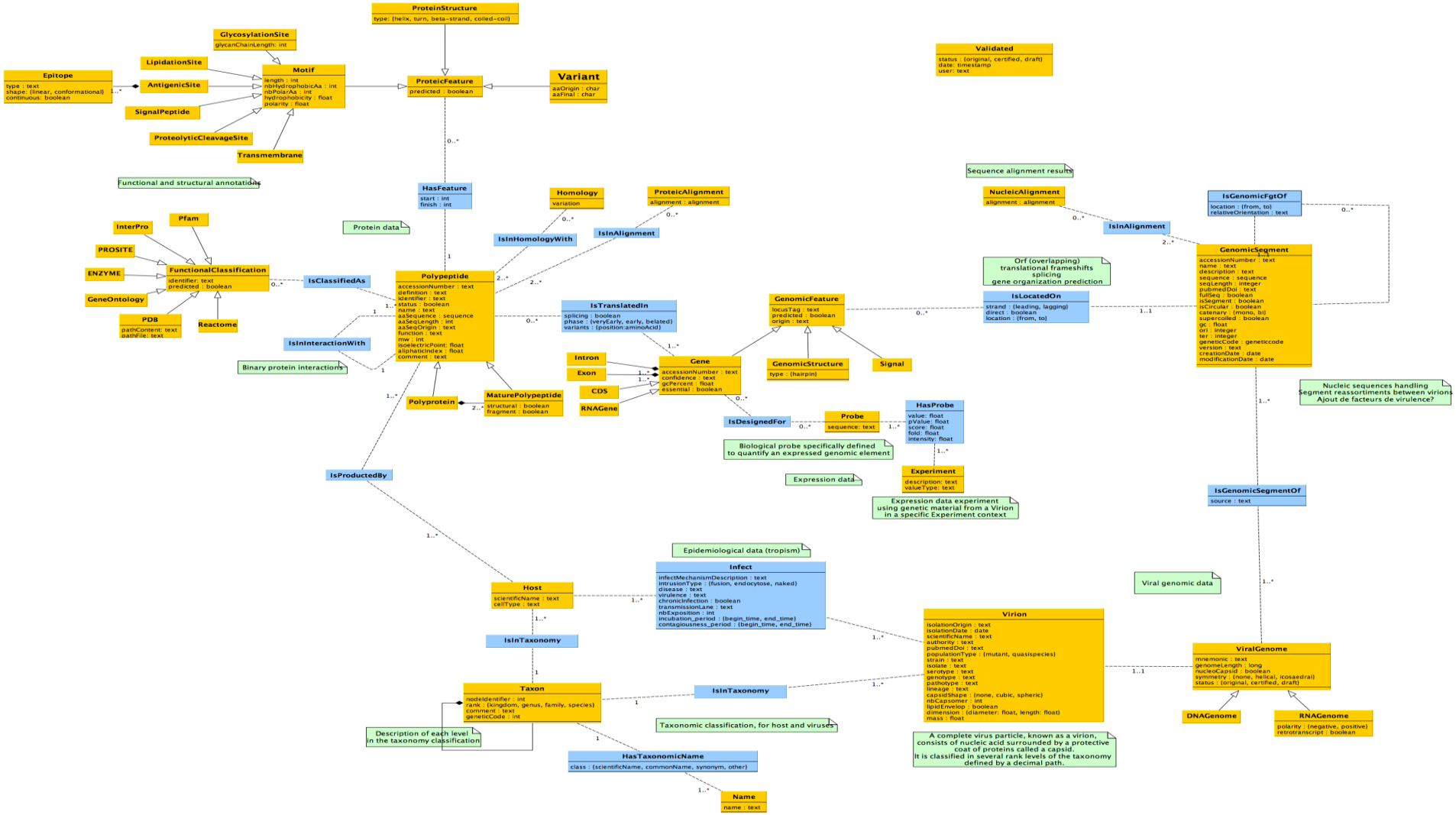
## Evolving biological data

- New technologies
- New problematics

# Data modeling via UML



# Data modeling via UML



# Advantages

- Intuitive (and graphical) UML-like representation of biological entities and of their relationships
- Formal modeling (*vs.* natural language): no ambiguity over the definition of entities and relationships
- An integrated data space as a large network where nodes are entities and edges are relationships
- Efficient support for data consistency checking
- Navigation and query facilities over the whole data space

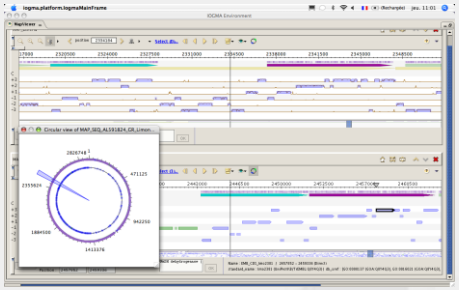
# Data modeling in software

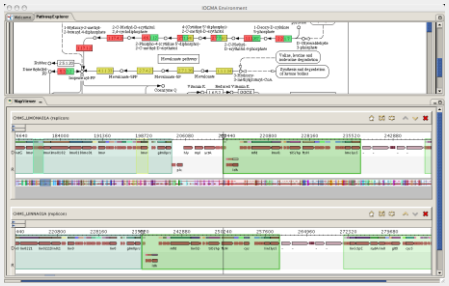
- Entities described as classes: types and subtypes
  - **Distinction between « sequence » and « replicon »**
- Relationships
  - **« Feature » is-located-on « sequence »**
- Methods described as classes
  - **Typed input and output**
- Typed input and ouput of methods
  - **Type checking: testing method adequacy for input data**
  - **Type assignment to output data**

# Data modeling in database

- MicroB: a relationnal database
  - **Interconnected genomic, proteic and metabolic reference data on more than 1500 microbial organisms**
- Overlapping schema with software schema
  - **More than 300 relations/tables**
  - **Easy data import and export from and back  to the software**

# An integrated bioinformatics platform

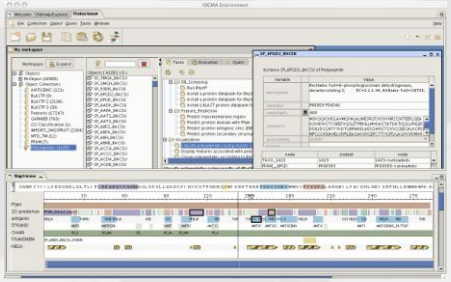




**Metabolic Pathway Builder**
Perform comparative genomics & metabolic analyses from annotation to analysis of relevant metabolic reactions & pathways

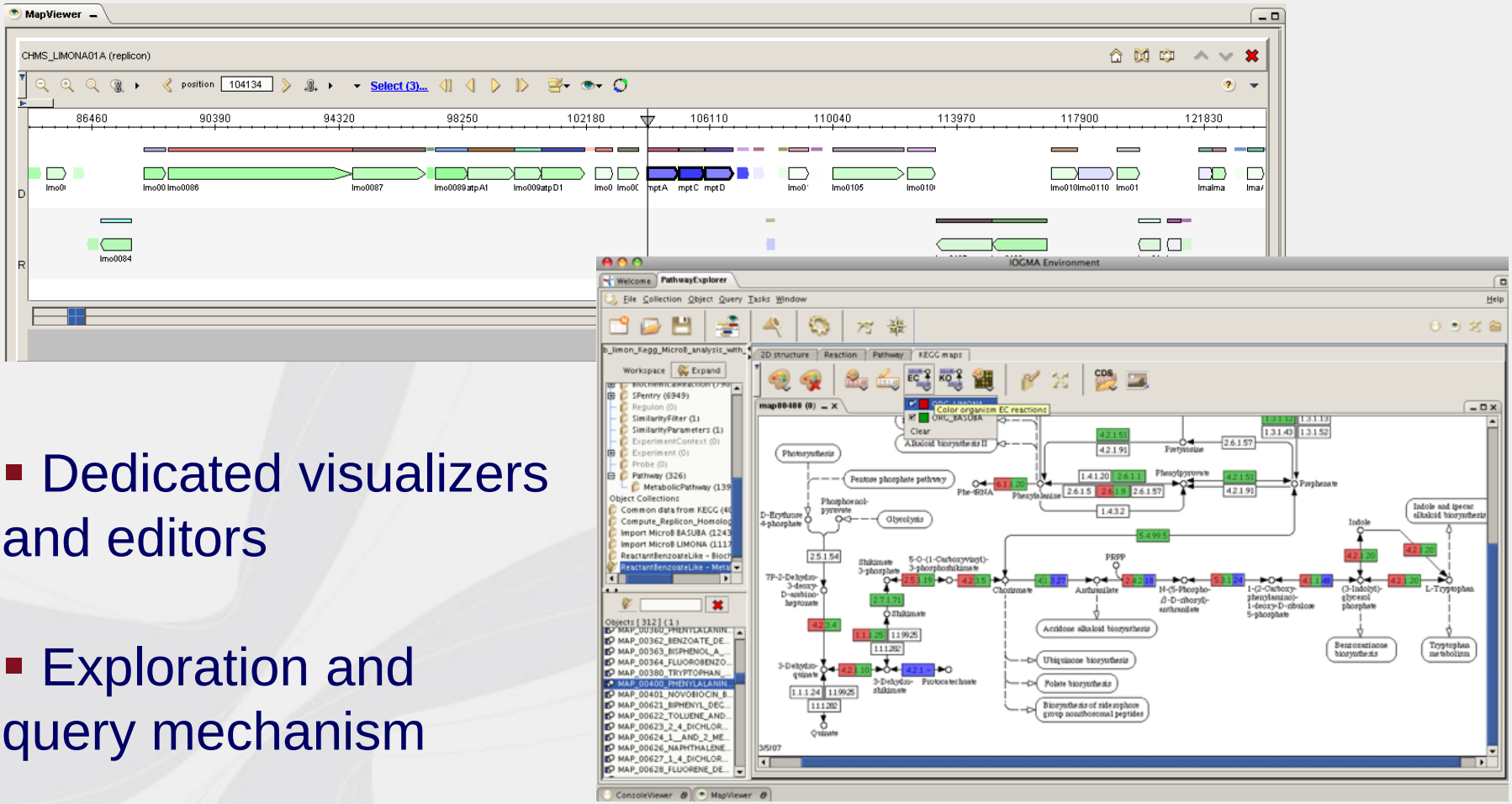
**MicroB database**
Connected genomic, proteic & metabolic data on 1500+ reference microorganisms

Integration of new annotated genomes



Genostar
Bioinformatics solutions

# An integrated bioinformatics platform




- Dedicated visualizers and editors
- Exploration and query mechanism

Genostar
Bioinformatics solutions

# Contacts

www.genostar.com

Francois.Rechenmann@genostar.com