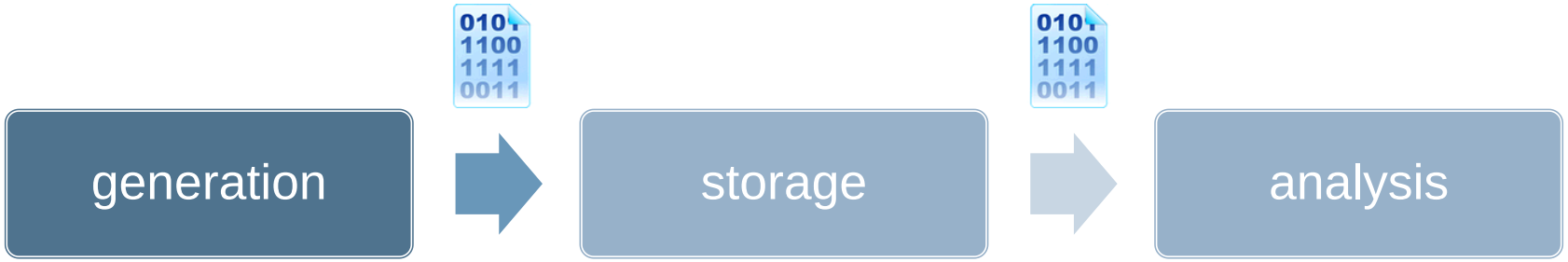
Gene Ontology based automation:

why it isn’t working

Genomics in a nutshell



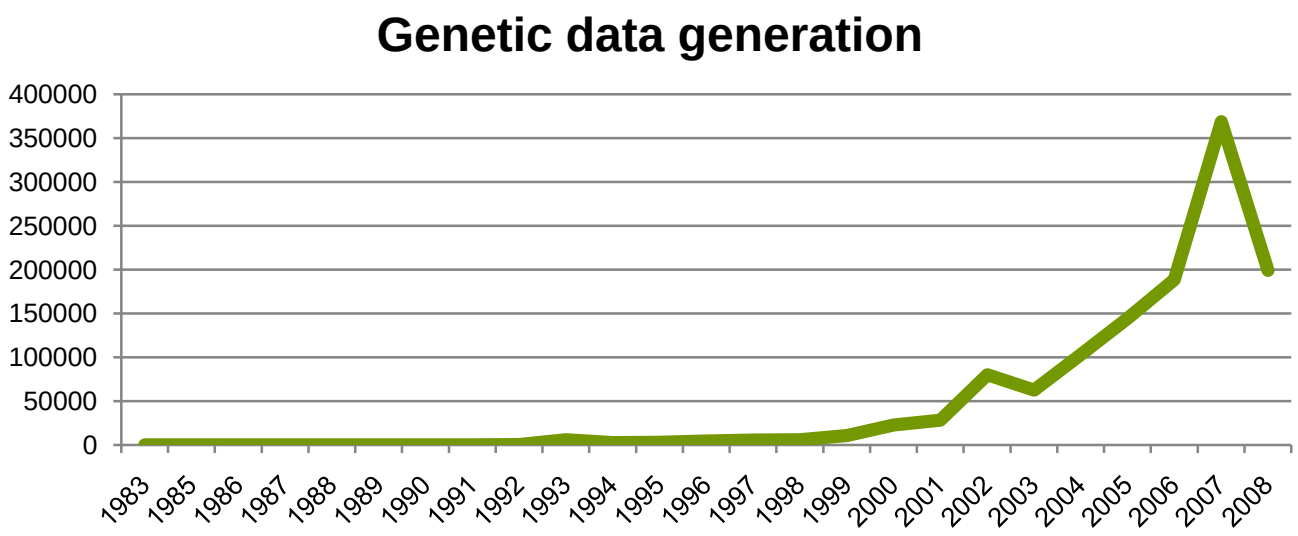
Current solutions

- **Lots** of **highly complex** data
- **Heterogeneous** data storage
- Unwanted **redundancy**
- **Inconsistencies** among the datasources

Problem: as of yet no holistic view!

Future prospects

The problems are only **getting worse**, considering that sequencing data volume is **expected to increase**



Current solutions

- Ontologies
  - Gene Ontology (GO)
    - "*It is unclear what kinds of reasoning are permissible on the basis of GO's hierarchies.*"
    - "*No procedures are offered by which GO can be validated.*"
- Ad-hoc databases
  - National Center for Biotechnology Information (NCBI)
  - The Human Gene Database (HGMD)
  - The Breast Cancer Information Core (BIC)
  - dbSNP
- Web Services
- Intelligent Agents

Conceptual modeling

Conceptual modeling

Understanding

- Represent / simulate reality

Design

- Model Driven Engineering

Communication

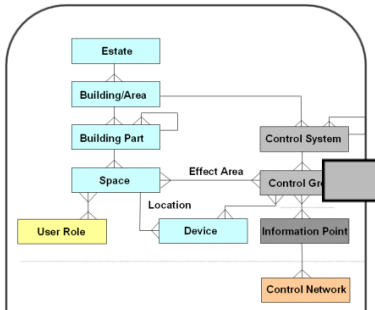
- Disambiguation
- Standardized data formats

- "*Life* as an Information System"
- Implicitly defines an Ontology
- Perform reasoning tasks

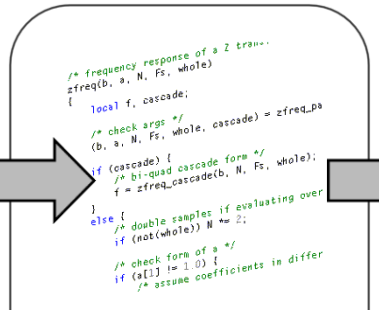
- Higher quality Genomic Information Systems
- Reuse of components

- More efficient information exchange

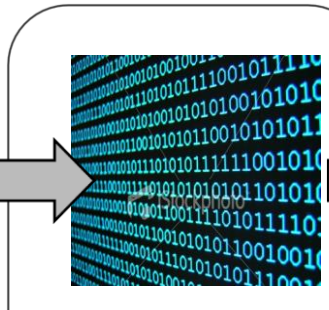
Life as an Information System



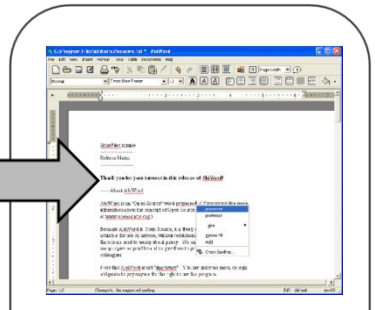
Abstract elements



Behaviour specification



Executable code



Implementation

Contribution

- **Solving data chaos**
  - Integrating fragmented data into one repository
    - A large, consistent set of genetic variation data
    - Data loading routines / algorithms
  - Creating higher quality Genomic Information Systems, faster!
- **Solving conceptual chaos**
  - Remove ambiguity
  - Bridge the semantic gap between domain experts and engineers
- **The Conceptual Schema as a tool for understanding**
  - Replacing/contributing to Gene Ontology