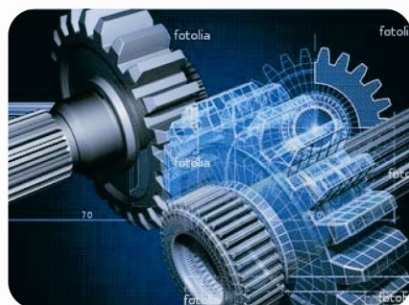
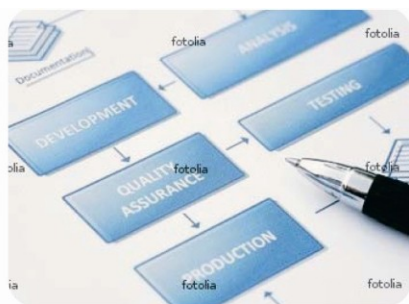




UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



Centro de Investigación en Métodos  
de Producción de Software



# Centro de Investigación ProS

## Future Doctoral Works

### I+D in Personalized Genomic Medicine



1. Problem Statement
2. Background
3. Goals
4. Projects
5. Ontology and GeIS
6. GeIS for Breast Cancer
7. Wiki-Genome
8. Doubts



## Genomic Data Chaos

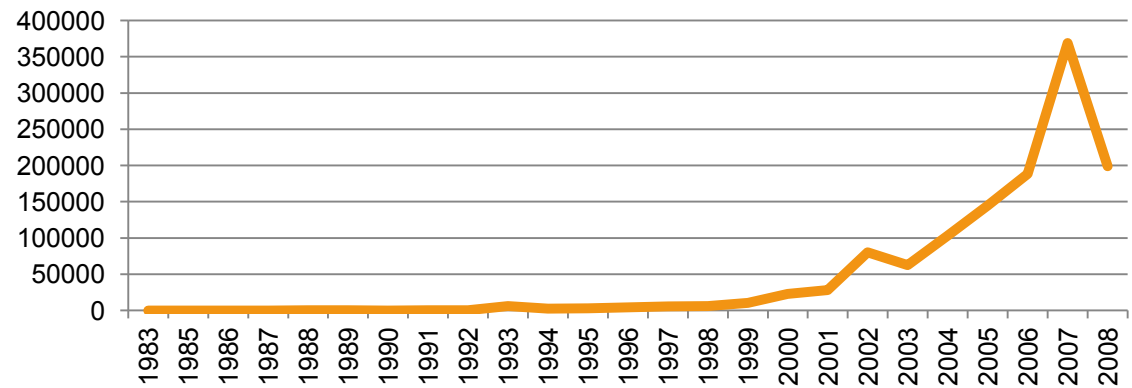




- **Lots of highly complex data**
- **Heterogeneous data storage**
- **Unwanted redundancy**
- **Inconsistencies among the data sources**

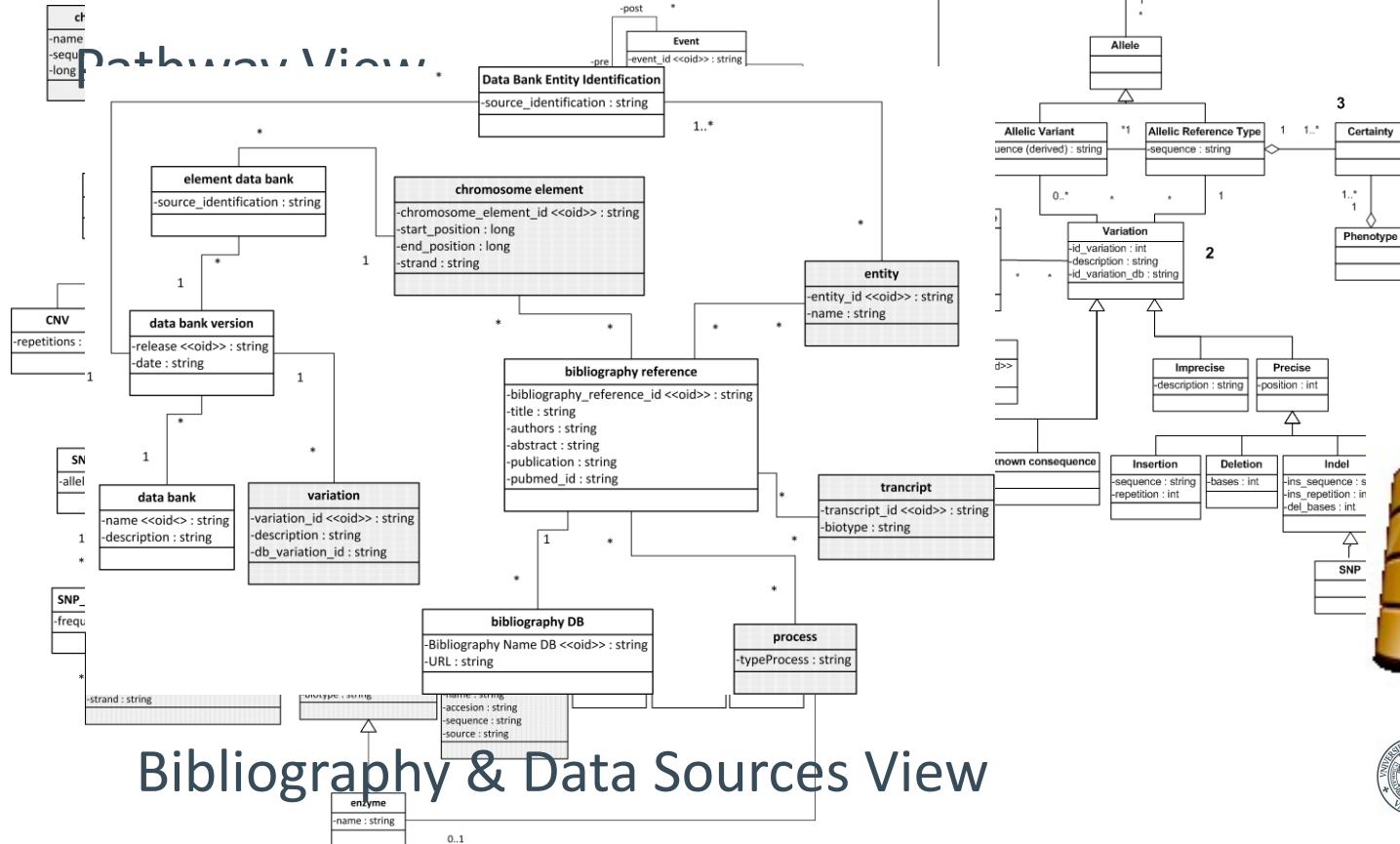
the  
holistic  
View

### Genetic data generation





- Conceptual Schema of Human Genome
- Human Genome Data Base
- Genome Information System





- Defining a genomic domain by using ontologies and conceptual models
- Designing and implementing a Genomic Information System (GeIS)
- Specifying of the GeIS for a particular disease (Breast Cancer)
- Designing and implementing a genomic data management system



- Human Genome Conceptual Modeling: A Ontological Framework for the Design and Implementation of Gels
- Design and Development of a Gels to Manage Breast Cancer Data
- Wiki-Genome: A Model-Driven Genomic Data Management Environment



# Human Genome Conceptual Modeling: An Ontological Framework for the Design and Implementation of GeIS

**Ainoha Martín Mayordomo**



- Main goal of the work: design and implementation of a GeiS:
  - Conceptual model
  - Ontology
  - Versioning system
  - Database
  - Loading strategies

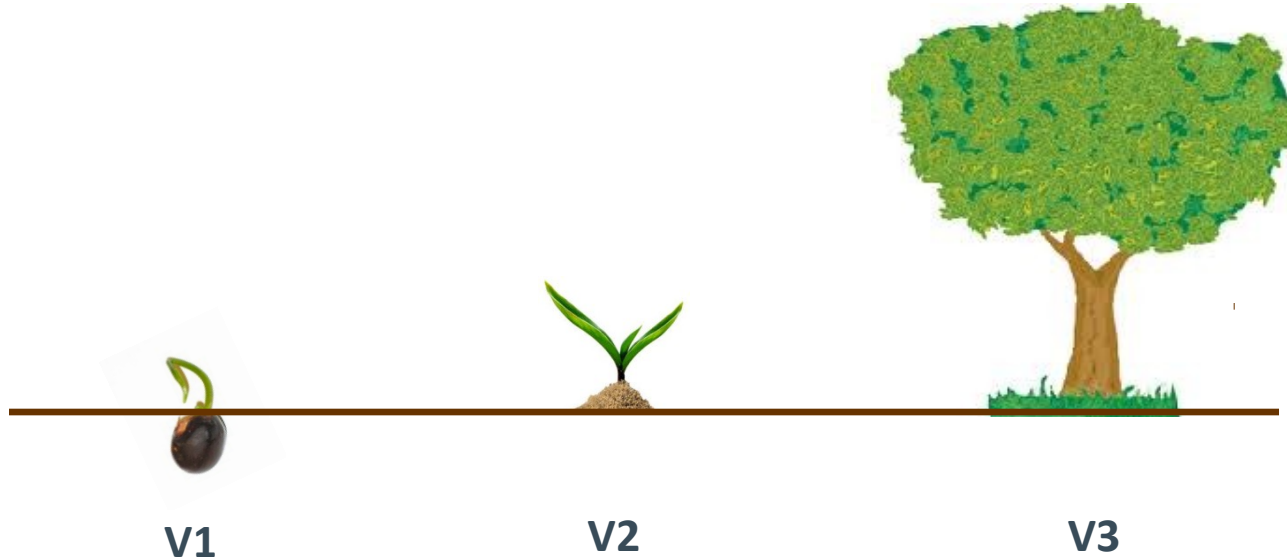


# Ontologies

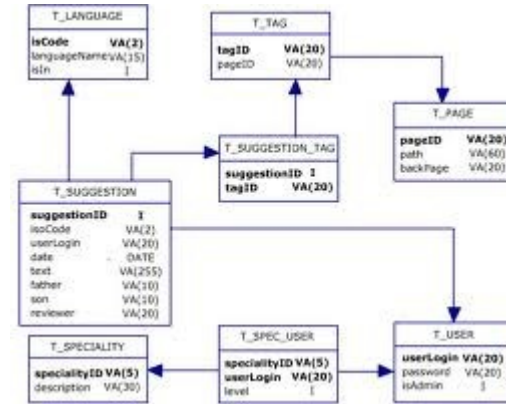
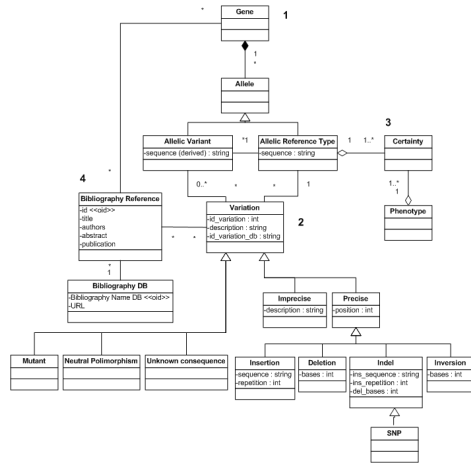
GFO

But... What about temporality?

- Provides a framework for building domain specific ontologies.
- Top-level ontology.
- Designed primarily for applications in medicine, biology and biomedical areas.
- New categories to define concepts (space, time...)



# Versioning System

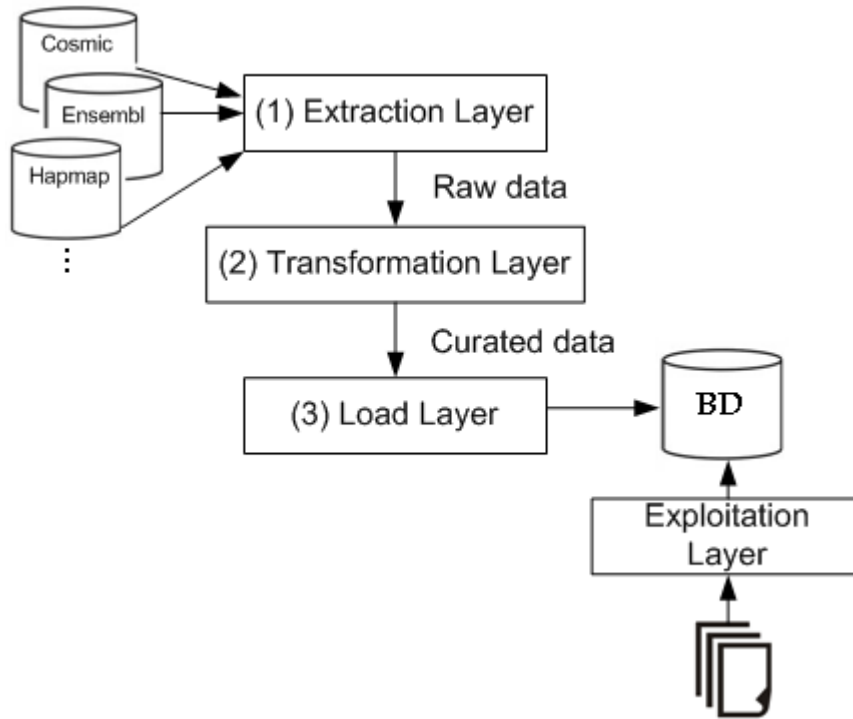


## Why use relational databases?

- Widely used
- Provides necessary technology to lead with large amount of data
- Non-redundant and structured organization of data
- Language SQL to retrieve and store information
- ...

## Why use Oracle?

- Most used
- Work in all platforms
- Good management users
- ...



↑ **independency**

↑ **flexibility**

↑ **scalability**

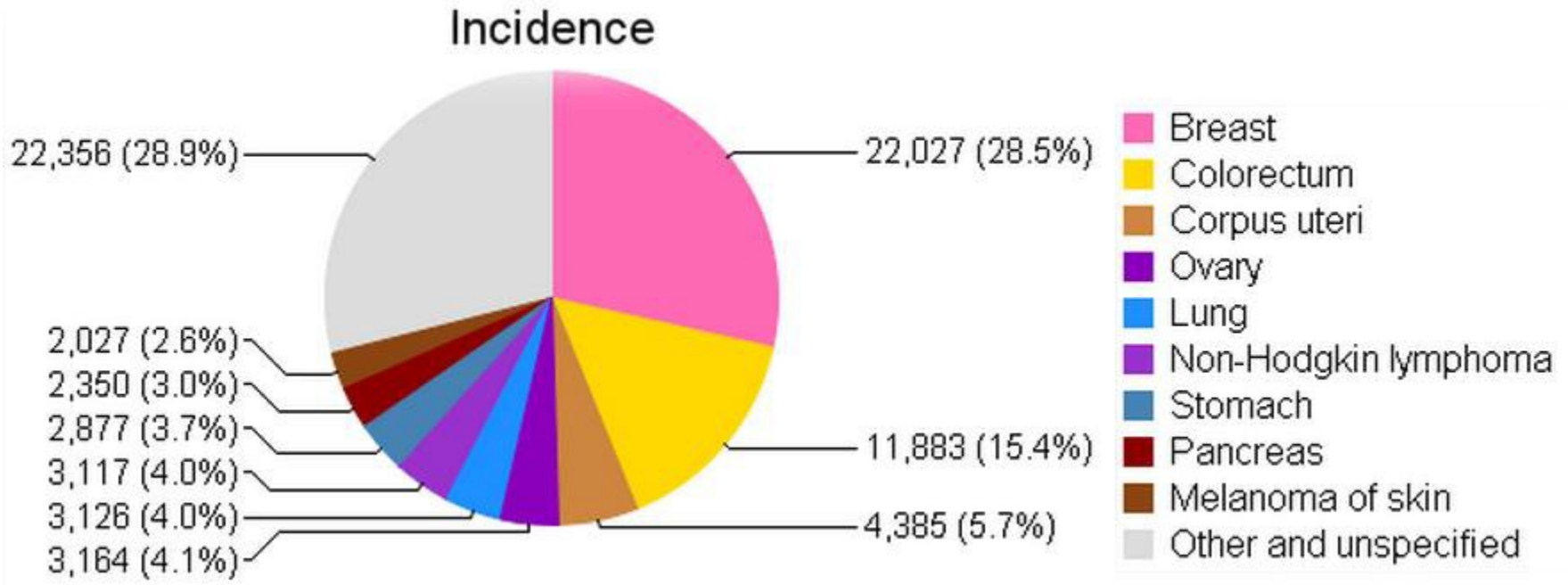
# Design and Development of a Genomic Information System to Manage Breast Cancer Data

Verónica Burriel Coll



## ■ Why Breast Cancer?

- Incidence of Breast Cancer in women society







- Data about Breast Cancer
  - The molecular characterization of BRCA1 and BRCA2 is complex
    - Large size
    - High density of mutations
  - Information is distributed by multiple databases
    - General genomic databases
    - Specific genomic databases
      - Genes
      - Diseases



- Proposed Methodology
  - List of relevant databases
  - Deep study of each database
  - Load the data about breast cancer contained
  - Extract the information not represented at CSHG
  - Extends the CSHG with new views
  - Update the HGDB following the CSHG's changes
  - Load the information on the updated database
  - Proof of concept



- List of relevant databases
  - Meetings with geneticists and clinicians
  - Classification of the databases
- Deep study of each database
  - Database structure
  - Contained information
  - Relevant information
  - Data formats
  - Data quality



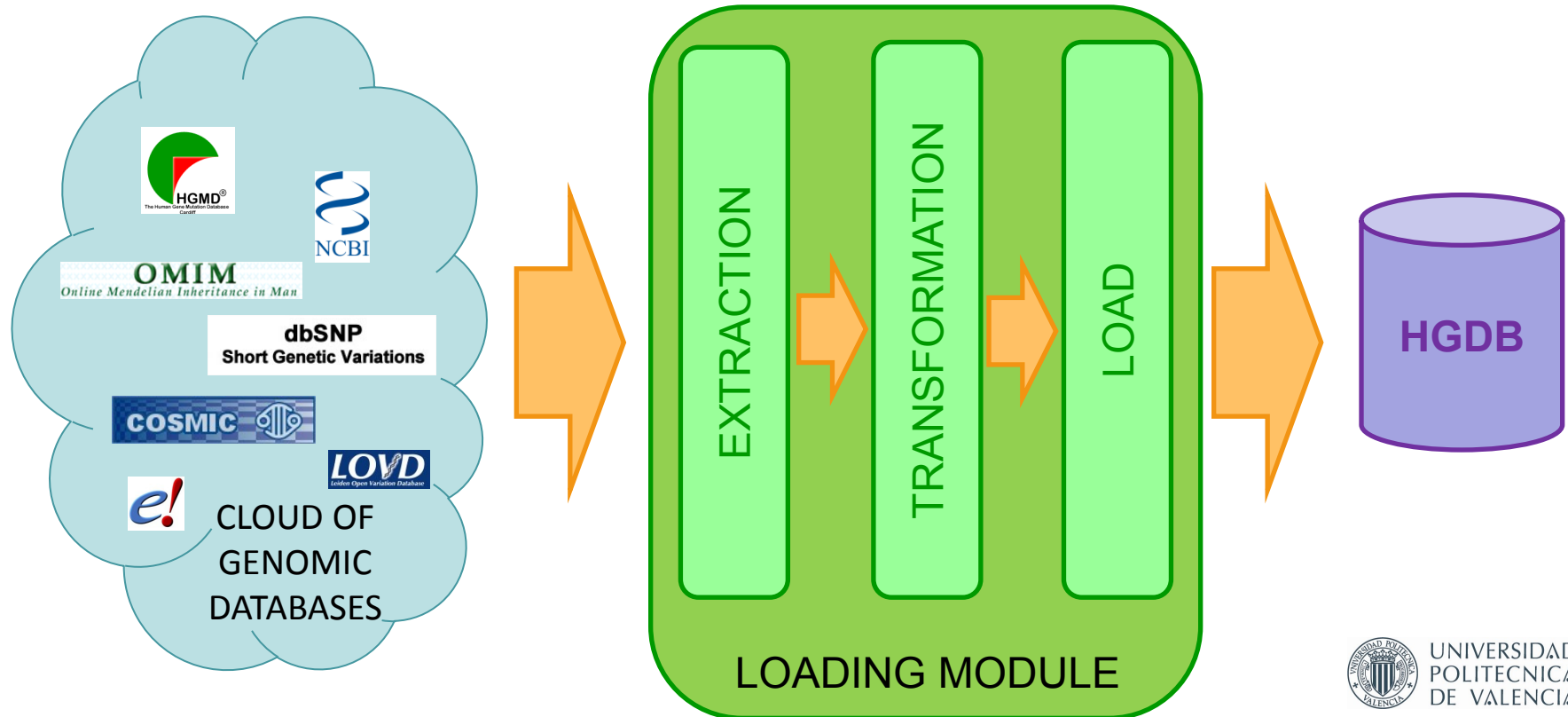
**dbSNP**  
Short Genetic Variations

**OMIM**  
Online Mendelian Inheritance in Man





- Load the contained data about breast cancer
  - Specific load modules for each database
  - Extract the relevant information



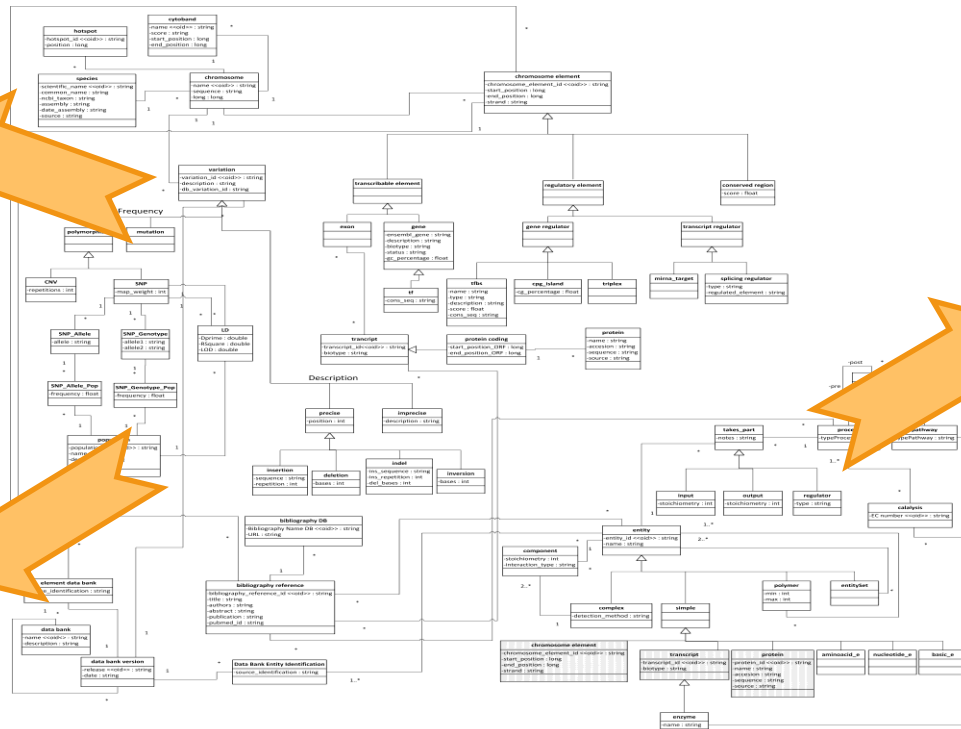
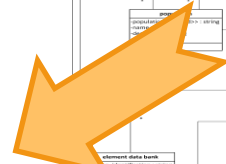


- Extract the information not represented in CSHG
  - Information which can't be inserted in HGDB
  - Select and study the environment
  
- Extends the CSHG with new views

**Phenotype**



**Drug**

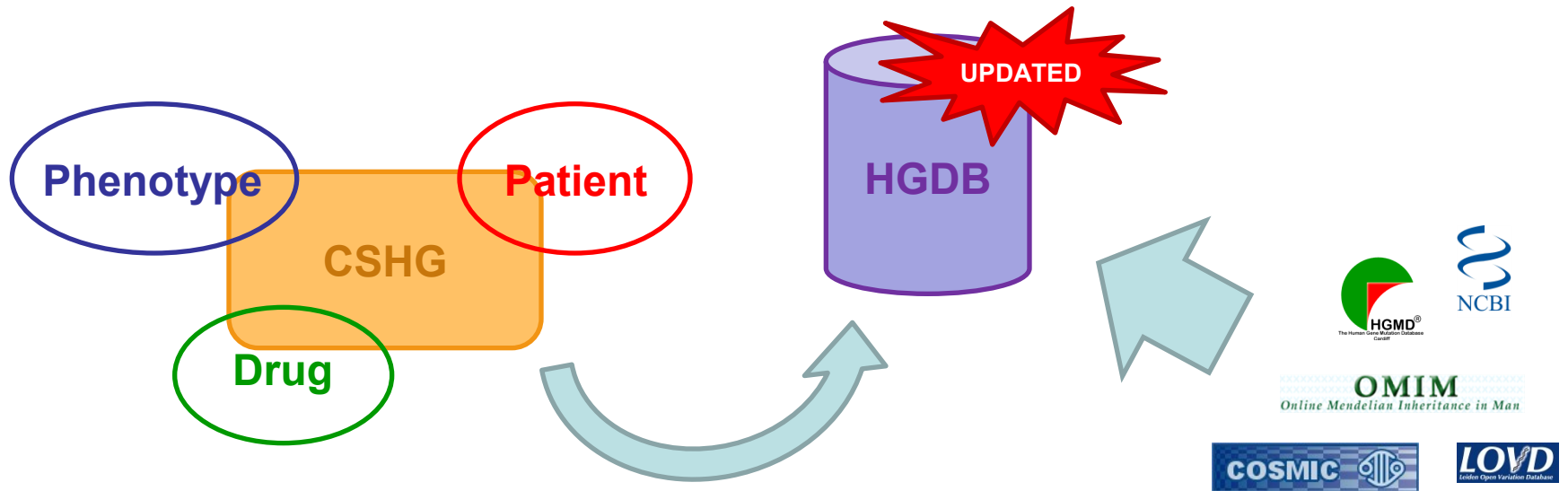


**Patient**



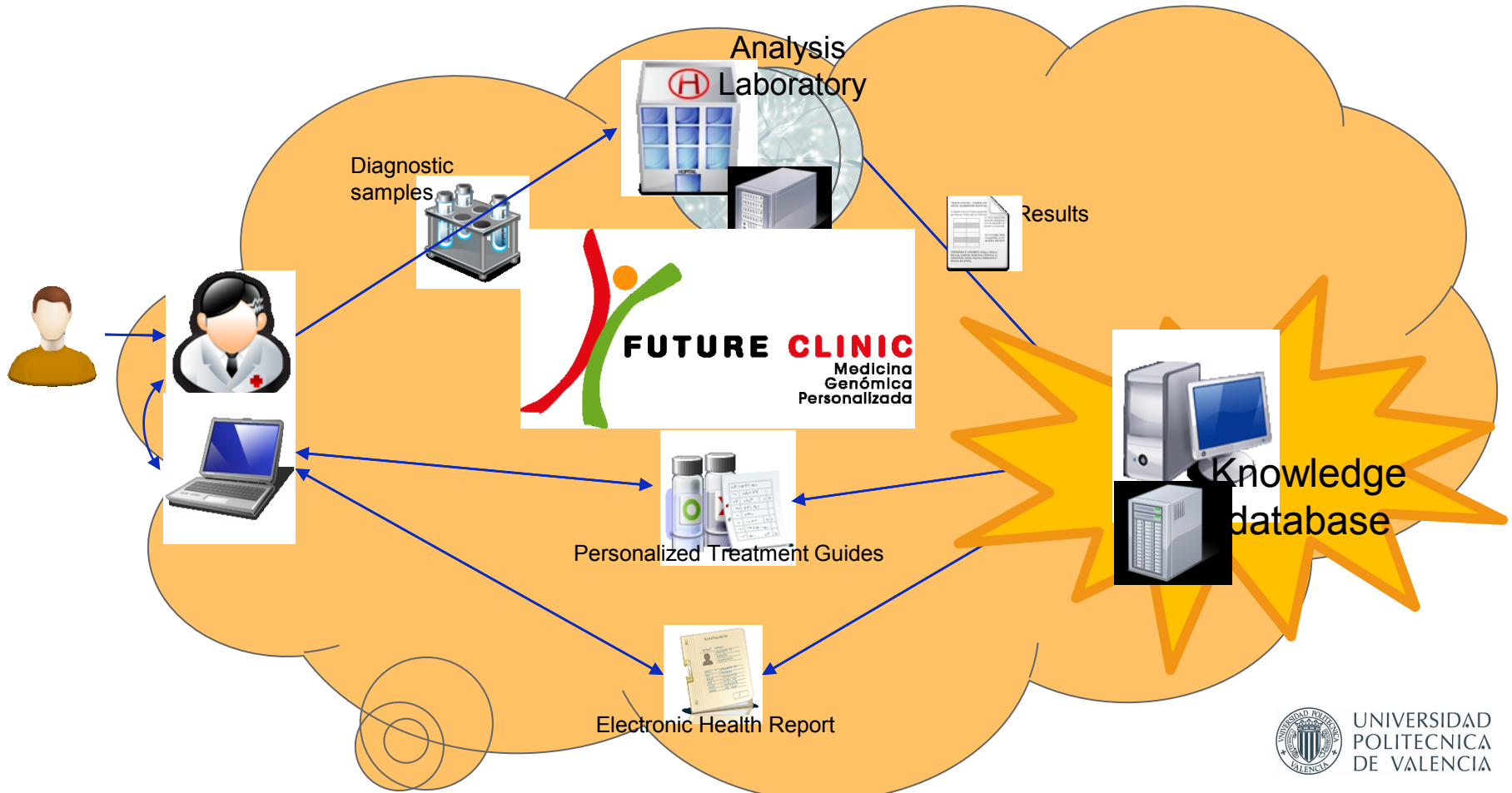


- Update the HGDB following the CSHG's changes
  - New HGDB's scheme extracted from CSHG
- Load the information on the updated database
  - Adapt the load modules





- Proof of concept: FutureClinic's Database
  - A reduced view adapted for the project purposes.





# Wiki-Genome: A Model-Driven Genomic Data Management Environment

Ana María Martínez Ferrandis

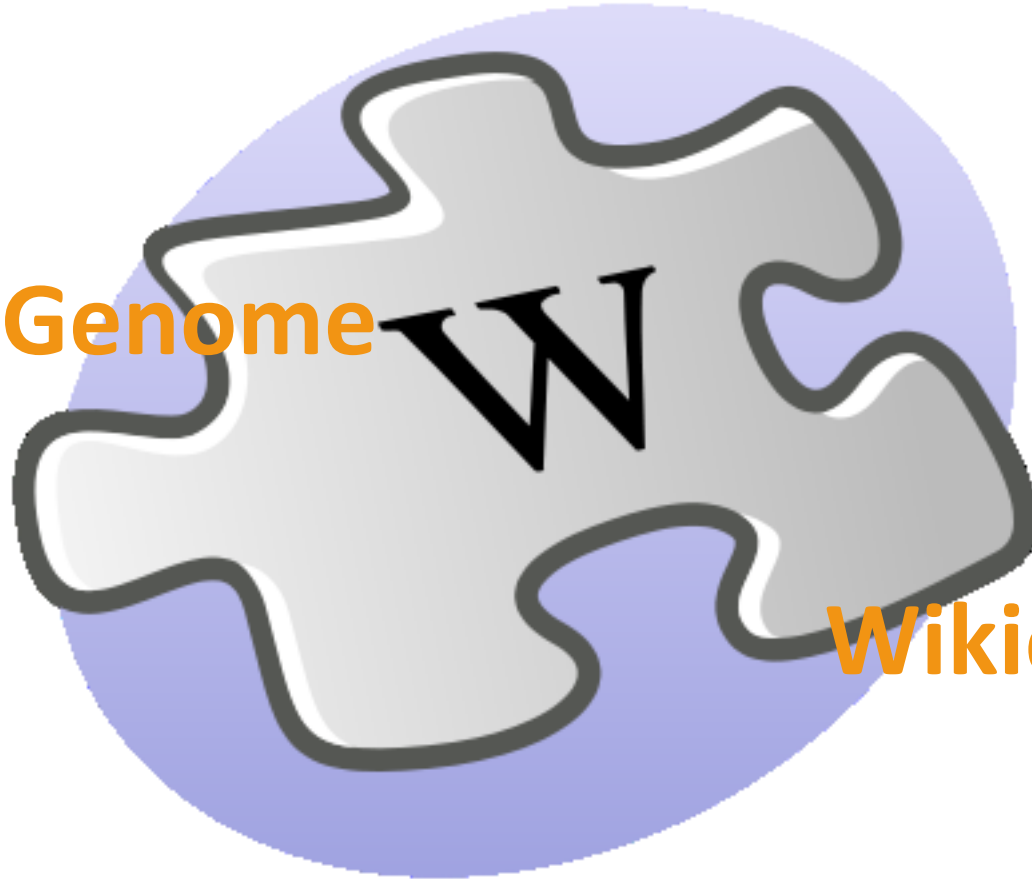


- ✓ Experts geneticists
- Information extraction
  - Documentation and bibliography
- Linking genes and phenotypes
- Introducing new data





**Wiki-Genome**



**Wikiomics**

# The beggining...



- Wikipedia ~~modeled~~ **consistent**
- Ex: Gene Wiki, Topsy, WikiGenes,  
WikiPathway, Wiki Professional...  
  - Without model or ontology **Coherent**





- BOWiki based on GFO - BO
- Background knowledge + reasoners
  - Limited automatic data review





- Overcome previous limitations (Wikipedia)
  - Solving coherent and consistency problems
    - Using Ontology based on GFO
  - Automatic data review + Curator
- 
- Implementation
    - OWL-DL
    - JENA
    - R2O
    - SPARQL

**Thank you!!!**

**Any doubt?**

