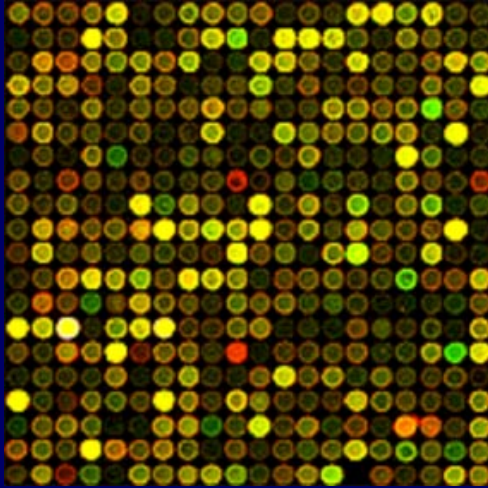


Using ontologies in clinico-genomic applications. The INFOBIOMED and ACGT experiences

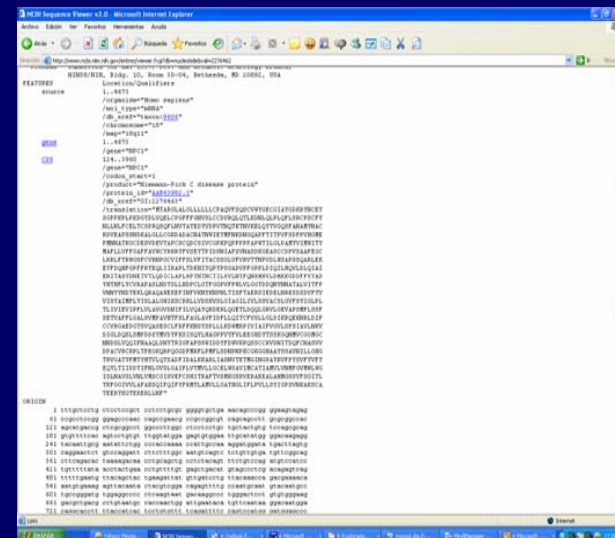
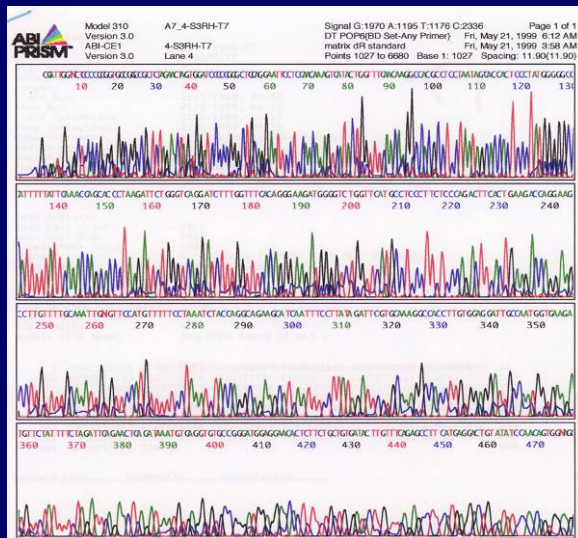
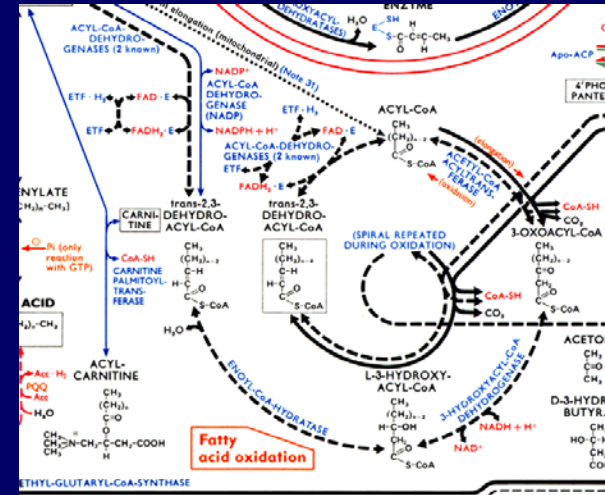
Dr. Miguel García Remesal
Biomedical Informatics Group
Artificial Intelligence Lab
Universidad Politecnica de Madrid, Spain (UPM)

mgarcia@infomed.dia.fi.upm.es

Biomedical Ontologies and DB integration



Knowledge extraction from remote, massively heterogeneous online sources and Integration with Clinical Information for research and patient care



Specific information sources

□ Michael Y. Galperin

The Molecular Biology Database Collection: 2007 update

Nucleic Acids Research Advance Access published on December 5, 2006

Nucl. Acids Res. 2007 35: D3-D4; doi:10.1093/nar/gkl1008

[\[Abstract\]](#) [\[FREE Full Text\]](#) [\[Print PDF\]](#) [\[Screen PDF\]](#) [\[Database Summaries\]](#) [\[Request Permissions\]](#)

Nucleic Acids Research

JOURNALS ONLINE

[HOME](#) [HELP](#) [FEEDBACK](#) [SUBSCRIPTIONS](#) [ARCHIVE](#) [SEARCH ARTICLES](#) [TABLE OF CONTENTS](#)

Database Categories List

- ▶ Nucleotide Sequence Databases
- ▶ RNA sequence databases
- ▶ Protein sequence databases
- ▶ Structure Databases
- ▶ Genomics Databases (non-vertebrate)
- ▶ Metabolic and Signaling Pathways
- ▶ Human and other Vertebrate Genomes
- ▶ Human Genes and Diseases
- ▶ Microarray Data and other Gene Expression Databases
- ▶ Proteomics Resources
- ▶ Other Molecular Biology Databases
- ▶ Organelle databases
- ▶ Plant databases
- ▶ Immunological databases

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

OUP is not responsible for the content of external internet sites

More than 950
public databases
with genes,
proteins and
diseases
(+thousands of
private databases)

Example: OXFORD (European Bioinformatics Institute)

The screenshot shows the EBI Services website in Microsoft Internet Explorer. The browser window title is "EBI Services - Microsoft Internet Explorer" and the address bar shows "http://www.ebi.ac.uk/services/". The website features a navigation tree on the left with categories: Databases, Toolbox, Submissions, Downloads, and Services Help. A "Services Overview" box is also present. The main content area is organized into several sections:

- Submissions**
 - AEdb
 - ArrayExpress via MIAMEExpress
 - EMBL via WEBIN
 - EMDep
 - MGTLA
 - PDB-AutoDep
 - UniProt via SPIN
 - Webin-Align
- Downloads**
 - EBI FTP Server
 - Help Files
 - Database Repository
 - Software Repository
- Toolbox**
 - Similarity & Homology**
 - Blast2 - ASD **NEW**
 - Blast2 - EVEC
 - Blast2 - NCBI
 - Blast2 - Parasite
 - Blast2 - WU
 - Fasta
 - Fasta - ASD **NEW**
 - Fasta - LGIC **NEW**
 - Fasta - Geno.Proteo.
 - MPSrch
 - more...
 - Prot. Function. Analysis**
 - CluSTR
 - GeneQuiz
 - InterProScan
 - more...
 - Proteomic Services**
 - Dasty
 - UniProt DAS
 - Sequence Analysis**
 - Align
 - ClustaWV **Updated**
 - GeneWise
 - PromoterWise
 - more...
 - Structural Analysis**
 - DALI
 - DeLLite **NEW**
 - Maxsprout
 - MSD Services
 - MSDFold
 - more...
 - Tools Miscellaneous**
 - EMBL Computational Services
 - Expression Profiler
 - NEWT
 - QuickGO
 - Readseq
 - Web Services
 - Whetizit **NEW**
 - more...
- Databases**
 - Database Browsing & Entry Retrieval via...**
 - BioMart
 - EMBL-SVA
 - Fetch Tools
 - Integr8 **NEW**
 - Query ArrayExpress
 - SRS
 - SRS3D
 - UniProt DAS **NEW**
 - UniProt Search **NEW**
 - WSDbfetch
 - Literature Databases**
 - MEDLINE
 - OMIM
 - Patent Abstracts
 - more...
 - Microarray Databases**
 - ArrayExpress
 - MIAME
 - Nucleotide Databases**
 - ASD
 - ATD **NEW**
 - EMBL-Bank
 - EMBL-CDS
 - Ensembl
 - Genome Reviews
 - MGTLA
 - more...
 - Protein Databases**
 - CSA
 - GOA
 - IntAct
 - IntEnz
 - InterPro
 - PANDIT
 - UniProtKB/Swiss-Prot
 - UniProtKB/TrEMBL
 - UniProt
 - more...
 - Proteomic Databases**
 - CHEBI
 - IntAct
 - IntEnz
 - more...
 - Structure Databases**
 - BioModels **NEW**
 - DALI
 - MSD
 - MSDchem
 - MSDPA

OXFORD: accessing a specific database

Standard Query Form - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos

Dirección <http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page=query+libList+OMIM> Ir Vinculos

EMBL-EBI
European Bioinformatics Institute

Quick Search Library Page **Query Form** Tools Results Projects Views Databanks **HELP**

Reset search **OMIM**

Search Options

Combine search terms with: & (AND)

Use wildcards

Get results of type: Entry

Result Display Options

View results using: OmimSimple

or

Create a view

Show 30 results per page

Tips

To do more advanced queries, use the [Extended Query Form](#).

Fields you can search Your search terms

In a single field, you can separate multiple values by &, |, ! **Search**

AllText

AllText

AllText

AllText

Create a view

Select the fields you want displayed in your view and choose the format

Choose 1 or more fields: Display As: Table List

ID
Keywords
MN
Text
SA
Reference
ClinicalSymptoms

Search

SRS Release 7.1.3.1 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. [Terms of Use](#) [Feedback & Support](#)

Listo Internet

Problems to solve

EBI Services - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos

Dirección: <http://www.ebi.ac.uk/services/>

Services Overview

- Databases
- Toolbox
- Submissions
- Downloads
- Services Help

Downloads

- EBI FTP Server
- Help Files
- Database Repository
- Software Repository

Submissions

- AEdb
- ArrayExpress via MIAMEExpress
- EMBL via WEBIN
- EMDep
- IMST/LLA
- FDB-AutoDep
- UniProt via SPIN
- Webin-Align

Toolbox

- Similarity & Homology**
 - Blast2 - ASD **NEW**
 - Blast2 - EVEC
 - Blast2 - NCBI
 - Blast2 - Parasite
 - Blast2 - VUJ
 - Fasta
 - Fasta - ASD **NEW**
 - Fasta - L GIC **NEW**
 - Fasta - Geno./Proteo.
 - MPsrch
 - more...
- Prot. Function. Analysis**
 - QUSTr
 - GeneQuiz
 - InterProScan
 - more...
- Proteomic Services**
 - Dasty
 - UniProt DAS
- Sequence Analysis**
 - Align
 - OustaW **Updated**
 - GeneMise
 - PromoterWise
 - more...
- Structural Analysis**
 - DALI
 - DalLite **NEW**
 - Maxsprout
 - NSD Services
 - NSDfold
 - more...
- Tools Miscellaneous**
 - EMBL Computational Services
 - Expression Profiler
 - NEWT
 - QuickGO
 - Readseq
 - Web Services
 - whatizit **NEW**
 - more...

Database Browsing & Entry Retrieval via...

- BioMart
- EMBL-SVA
- Fetch Tools
- Integr8 **NEW**
- Query ArrayExpress
- SRS
- SRS3D
- UniProt DAS **NEW**
- UniProt Search **NEW**
- vSDbfetch

Literature Databases

- MEDLINE
- OMIM
- Patent Abstracts
- more...

Microarray Databases

- ArrayExpress
- MIAME

Nucleotide Databases

- ASD
- ATD **NEW**
- EMBL-Bank
- EMBL-CDS
- Ensembl
- Genome Reviews
- IMGT/LLA
- more...

Protein Databases

- CSA
- COA
- IntAct
- IntEnz
- InterPro
- PANDIT
- UniProtKB/Swiss-Prot
- UniProtKB/TREMBL
- UniProt
- more...

Proteomic Databases

- CHEBI
- IntAct
- IntEnz
- more...

Structure Databases

- BioModels **NEW**
- DALI
- MSD
- MSDchem
- NCPA

WHAT'S 2can?

This logo is a link to a relevant section in the EBI's new bioinformatics educational website, "2can Bioinformatics".

Live EBI News Feed

XMLRSS 2

- There are no effective searching tools, that can organize and combine different resources, adapted to each user's needs

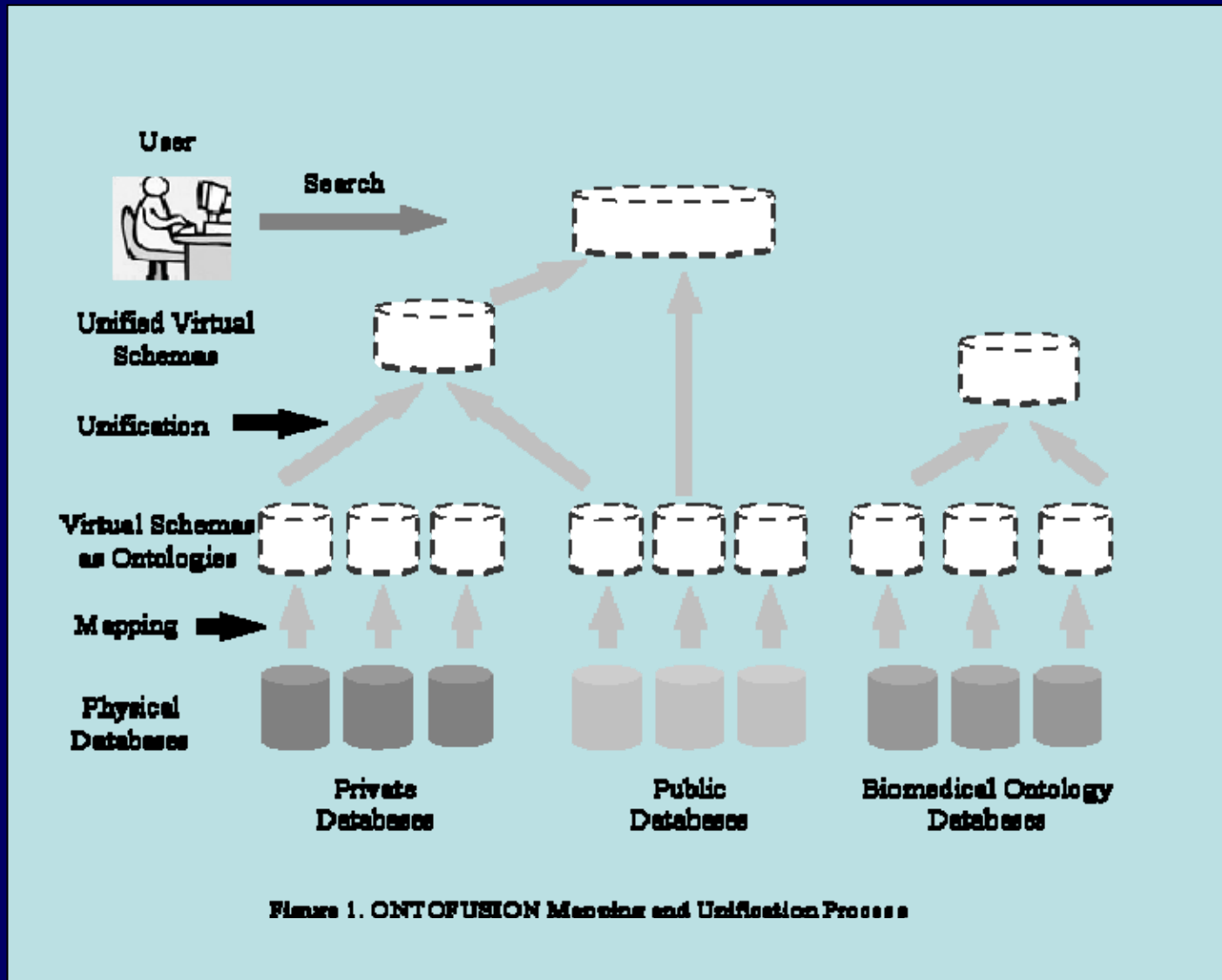
Integration needs

- Thousands of heterogeneous databases:
 - Distributed nature of sources
 - Different Platforms & Operating Systems
 - Database Management Systems
 - Different Semantics

Developing ontologies in biomedicine

- The Biomedical perspective from a strict definition of ontology: “An ontology is a formal, explicit specification of a shared conceptualization” is difficult to achieve:
 - *Formal*: difficult to express medical statements in pure logical form
 - *Explicit*: highly complex (now) to make explicit all the components of the medical reasoning processes due to cognitive causes
 - *Shared (consensus)*: high variability in medical practice and knowledge gaps make many times difficult to achieve agreements
 - Continuous new findings immediately modify wide consensus and the meaning of some diseases and facts (e.g., gastric ulcer)
 - “Consensuated” medical documents (e.g., practice guidelines) plagued with inconsistencies
 - *Conceptualization*: many past failed attempts in biomedicine

Using ontologies to solve semantics inconsistencies within INFOBIOMED



Mapping Window

Physical Schema Mapping Virtual Schema Domain Ontology Finish Help

Physical Schema

- Physical Schema
- attribute
- c_appears_in
- c_has_attributes
- c_has_sna
- co_occurrence
- comes_from
- concept
- context_attr
- context_attr_go
- context_attr_hi7
- context_attr_iod10am
- context_attr_iod9cm
- context_attr_msh
- context_attr_smi
- definition
- expression
- has
- has_context_relation

Virtual Schema

- Virtual Schema
- UMLS_Concept
- UMLS_Definition
 - UMLS_Definition.CUI
 - UMLS_Definition.Definition
 - UMLS_Definition.is_a_definition_of.UMLS
- UMLS_Codification
- UMLS_Semantic_Type
- UMLS_Synonym
- UMLS_Relation
- UMLS_Source

Domain Ontology

- Domain Ontology
- Synonym
 - GO_Gene_Product_Synonym
 - GO_Synonym
 - UMLS_Synonym
- GO_External_DB_Ref
- UMLS_Relation
- UMLS_Semantic_Type
- UMLS_Source
- GO_Evidence
- Definition
 - UMLS_Definition
 - UMLS_Definition.Definition
 - UMLS_Definition.CUI
 - UMLS_Definition.is_a_definition_of
 - GO_Definition
- Concept
 - GO_Gene_Product

Classes

- Mapping
- UMLS_Concept
- UMLS_Source
- UMLS_Definition
- UMLS_Semantic_Type
- UMLS_Relation
- UMLS_Codification
- UMLS_Synonym

Class Attributes

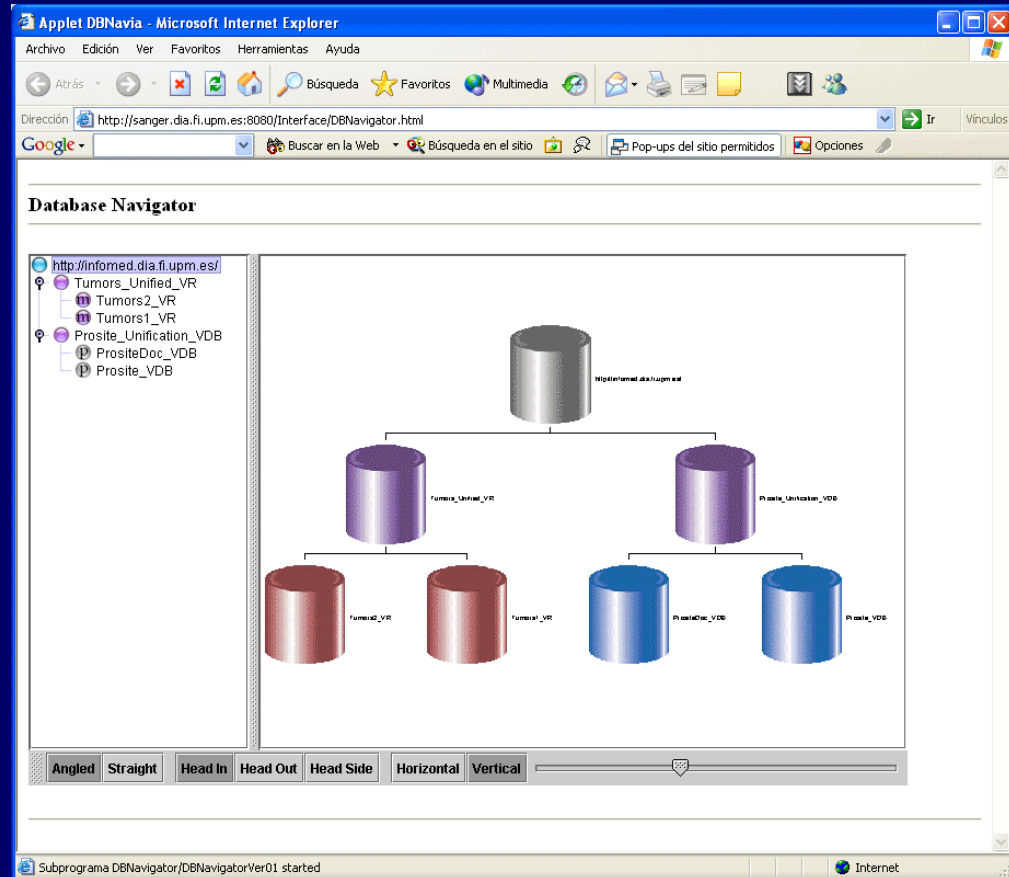
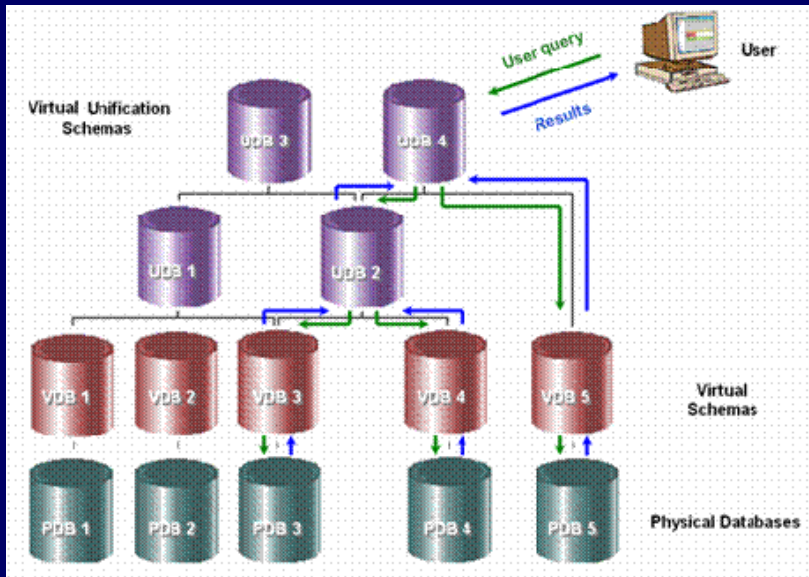
- UMLS_Definition
 - UMLS_Definition.CUI
 - is_defined_by.CONCEPT_CUI
 - UMLS_Definition.Definition
 - definition.DEFINITION_DEF

Class Relations

UMLS_Definition

- UMLS_Definition.is_a_definition_of.UMLS_Concept
- is_defined_by TO str2cui2lui
 - joinCondition
 - is_defined_by.CONCEPT_CUI
 - str2cui2lui.CONCEPT_CUI

ONTOFUSION Mapping Tool



(diagram)

Database navigator

(actual)

Ontonavigator JSP

http://infomed.dia.fi.upm.es#umls_VDB2

Open Close GetScale

umls_VDB2

- UMLS_Concept
- UMLS_Codification
 - Context_Code
 - SUI
 - Context_String
 - Hierarchical_code
 - CUI
 - is_a_codification_of.UML
- UMLS_Semantic_Type
- UMLS_Definition
- UMLS_Source
- UMLS_Relation

Conc... UMLS Codification UMLS S

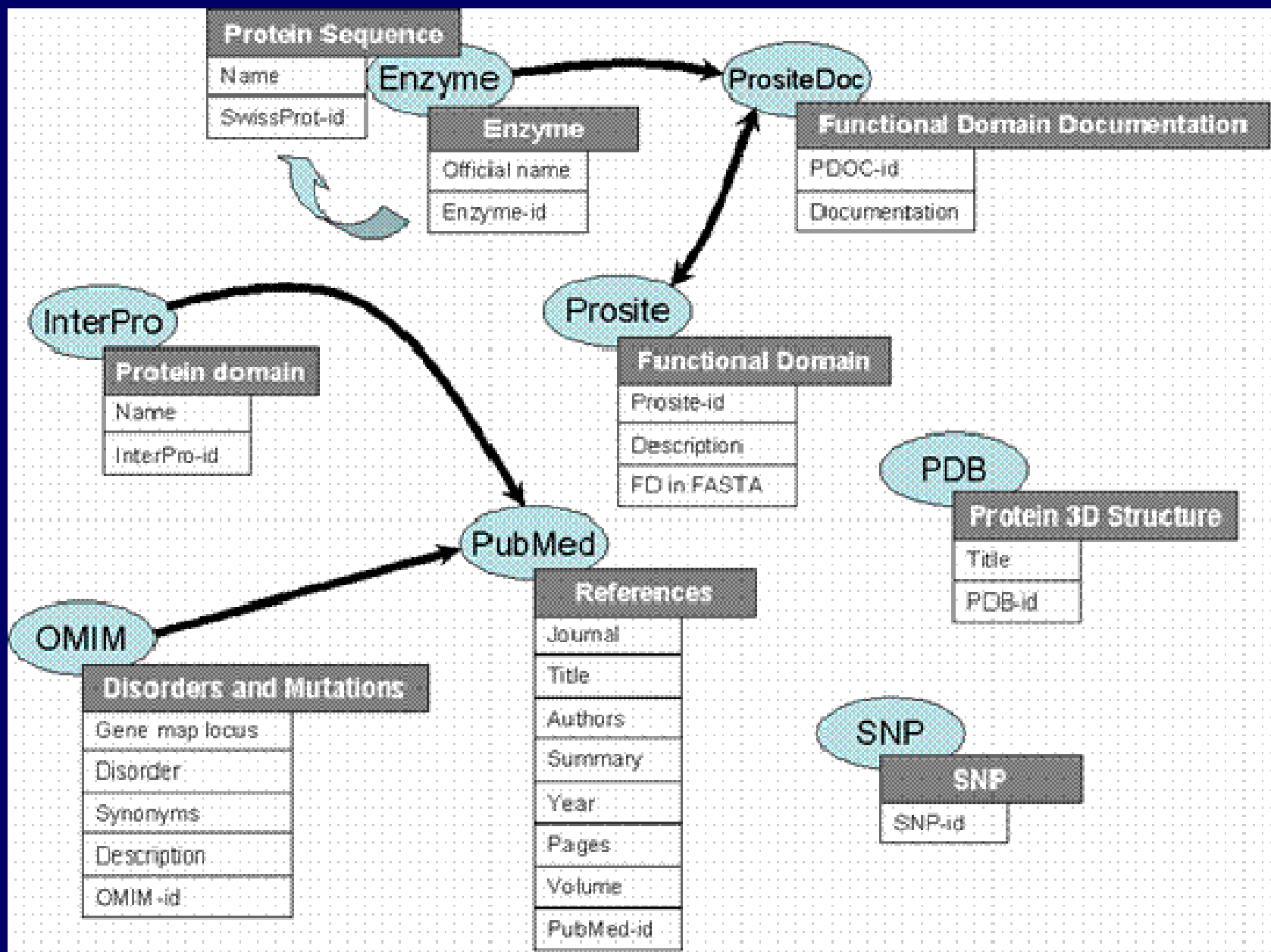
<input checked="" type="checkbox"/> Context_Code	<input type="radio"/> Contai...	A...	
	<input checked="" type="radio"/> Is	OR	
<input checked="" type="checkbox"/> SUI	<input type="radio"/> Contai...	A...	
	<input checked="" type="radio"/> Is	OR	
<input checked="" type="checkbox"/> Context_String	<input type="radio"/> Contai...	A...	
	<input checked="" type="radio"/> Is	OR	
<input checked="" type="checkbox"/> Hierarchical_code	<input type="radio"/> Contai...	A...	
	<input checked="" type="radio"/> Is	OR	
<input checked="" type="checkbox"/> CUI	<input type="radio"/> Contai...	A...	
	<input checked="" type="radio"/> Is	OR	
<input type="checkbox"/> is_a_codification_of.UMLS_Concept			<input type="button" value="→"/>

submit

Angled Straight Head In Head Out Head Side Horizontal Vertical

Unification of public gene, protein and disease DBs

- In many cases, the unification might not be actually practical, since none concept appears in several DBs
- In this case, the unification process establishes links between different DBs (e.g., to relate instances of concepts from one DB to instances of another concept from another DB)



ACGT (Advancing Clinico Genomic Trials on Cancer)

- A European-funded Integrated Project to develop new models and architectures for supporting clinico-genomic trials
- Proposal: A Grid-based architecture, linking clinico genomic data from remote sources and different tools
 - Semantic Mediator
 - Oncosimulator
 - Distributed Data Mining
 - ...

Vision and opportunities

Many real problems waiting to be solved:

- Design of biobanks,
- Unification of SNP databases,
- Use of genomic information for primary care,
- Use of clinical data in drug discovery,
- Use of ontologies for developing new approaches for training physicians on “omics” areas, improving data mining and searching

Thanks!