Extracting multilevel information in biomedicine

(or going open, nano and abroad)

Victor Maojo

Professor and Director
GIB-UPM
Biomedical Informatics Group
Universidad Politecnica de Madrid
vmaojo@fi.upm.es

BIOMEDICAL INFORMATICS GROUP (GIB). DIRECTOR: VICTOR MAOJO

http://www.gib.fi.upm.es/

Main research areas

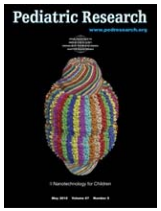
- Nanoinformatics and Nanomedicine
- Integration of distributed clinico-genomic databases
- Image processing and analysis
- KDD
- Biomedical ontologies
- Medical protocols
- Artificial Intelligence in Medicine (data, text and Web mining, expert systems, etc)

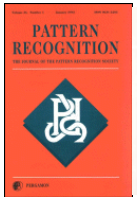
EC projects (only in the last two years):

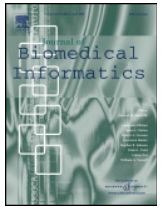
- 8 projects: ACGT, ACTION-Grid (Coordinator), DICODE, P-medicine, INBIOMEDVision,
- INTEGRATE, AFRICA-Build (coordinator), EUREKA (2012)

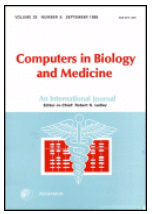
- External collaborations (outside the EU)

Publications (selection):





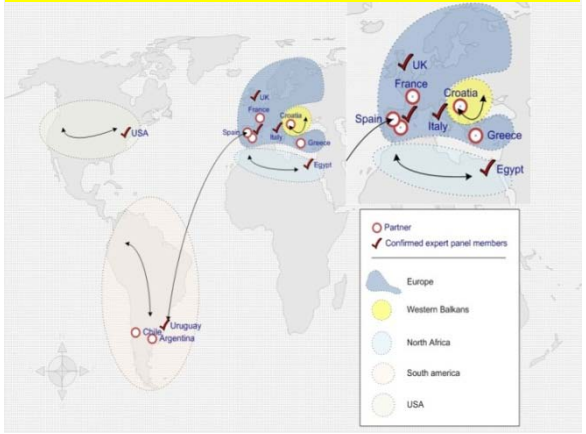


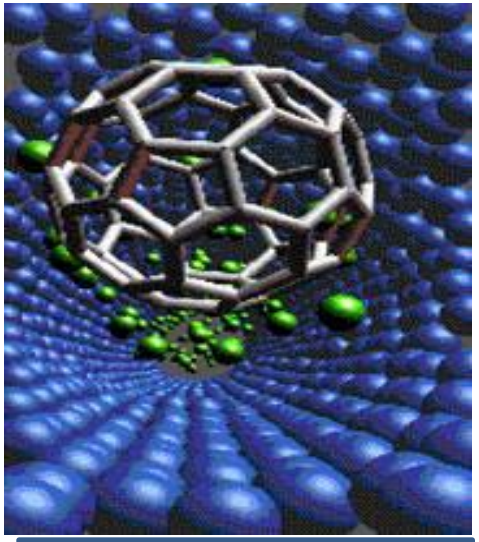



Three topics, from current initiatives at the GIB-UPM, but (surely) with a broader interest

Going open

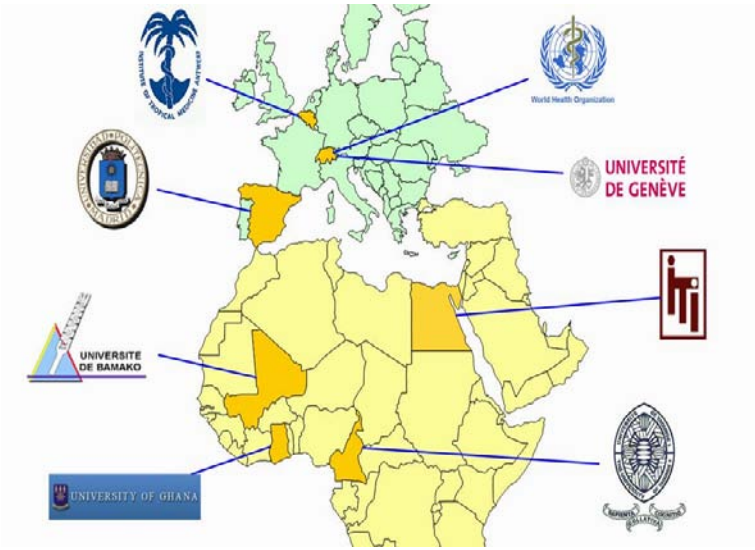
A global vision of biomedical software





Going nano

Extending Biomedical Informatics : Nanoinformatics

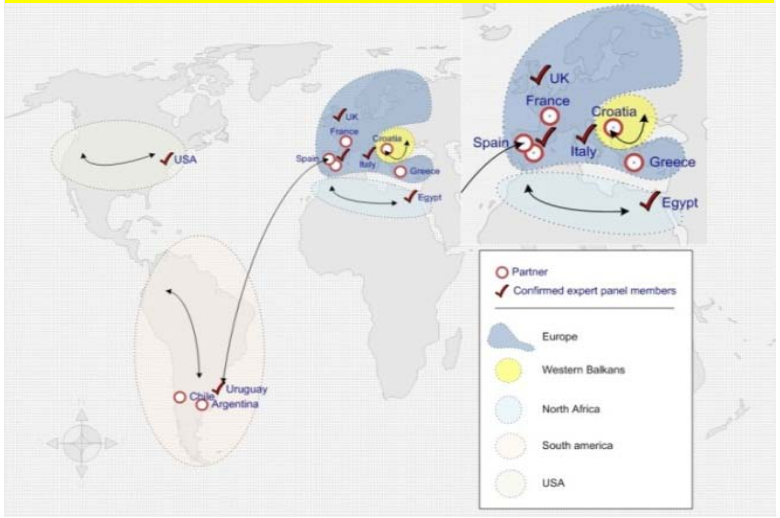


Going South

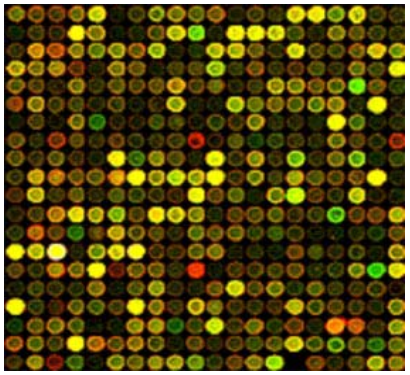
Using Biomedical Informatics to improve medical research and care in Africa

Going open

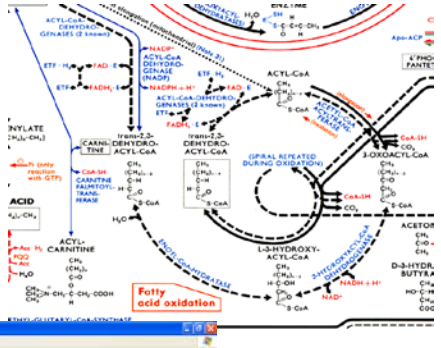
A global vision of biomedical software

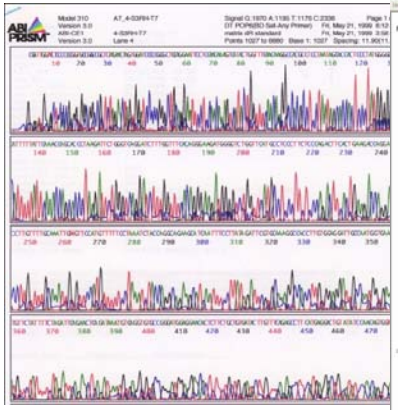


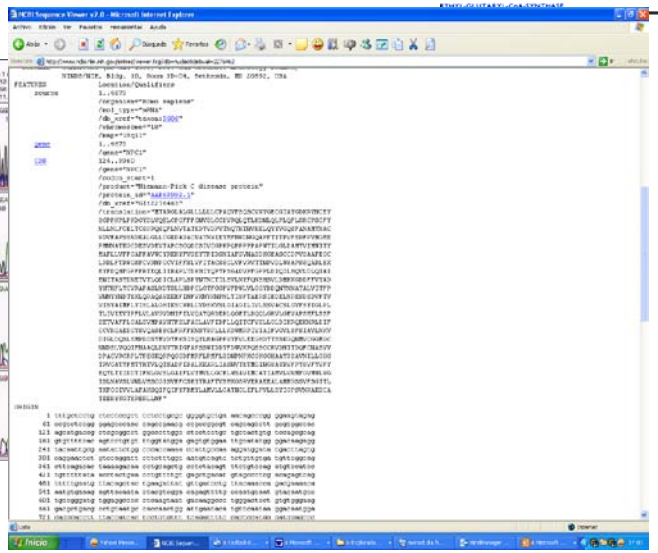
Postgenomic challenge (2001) : To extract knowledge from massive, heterogeneous (biological and clinical) information for genomic and personalized medicine



DNA sequences, metabolic pathways, images, etc

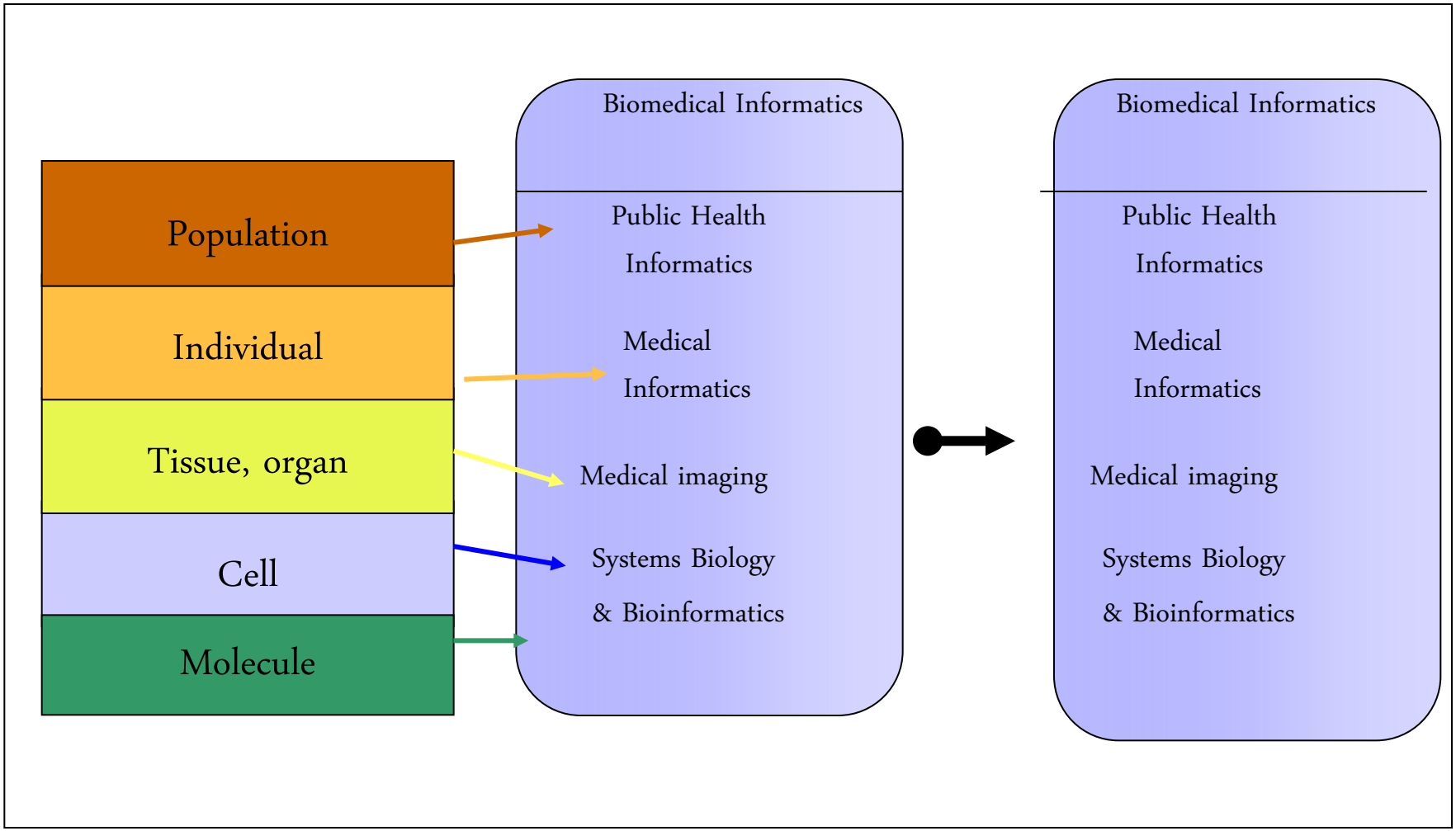


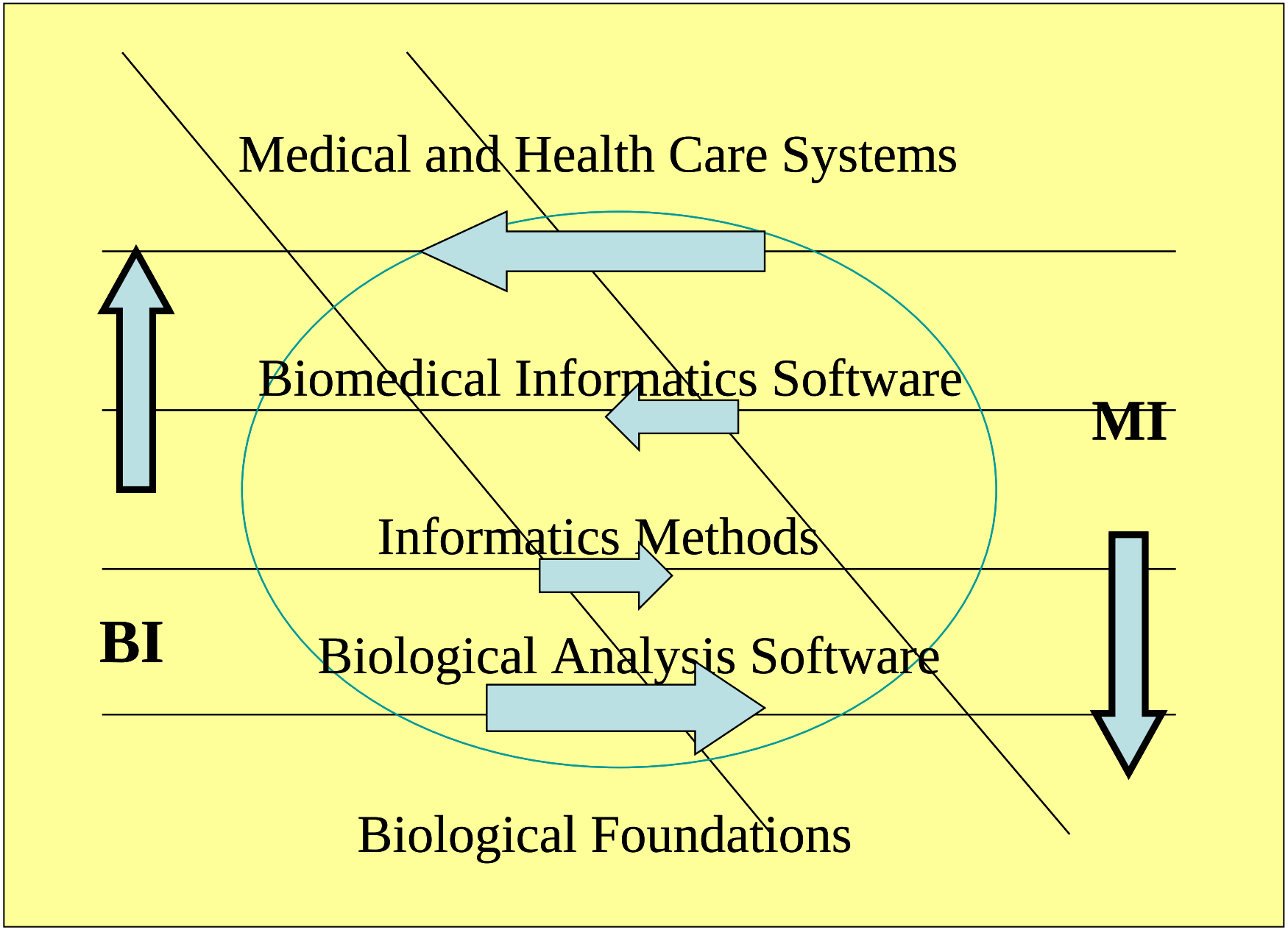




Cork, Sept 2011, Victor Maojo

(SCOPE) From anatomy (macro) towards microlevels (systems biology) and Nanomedicine

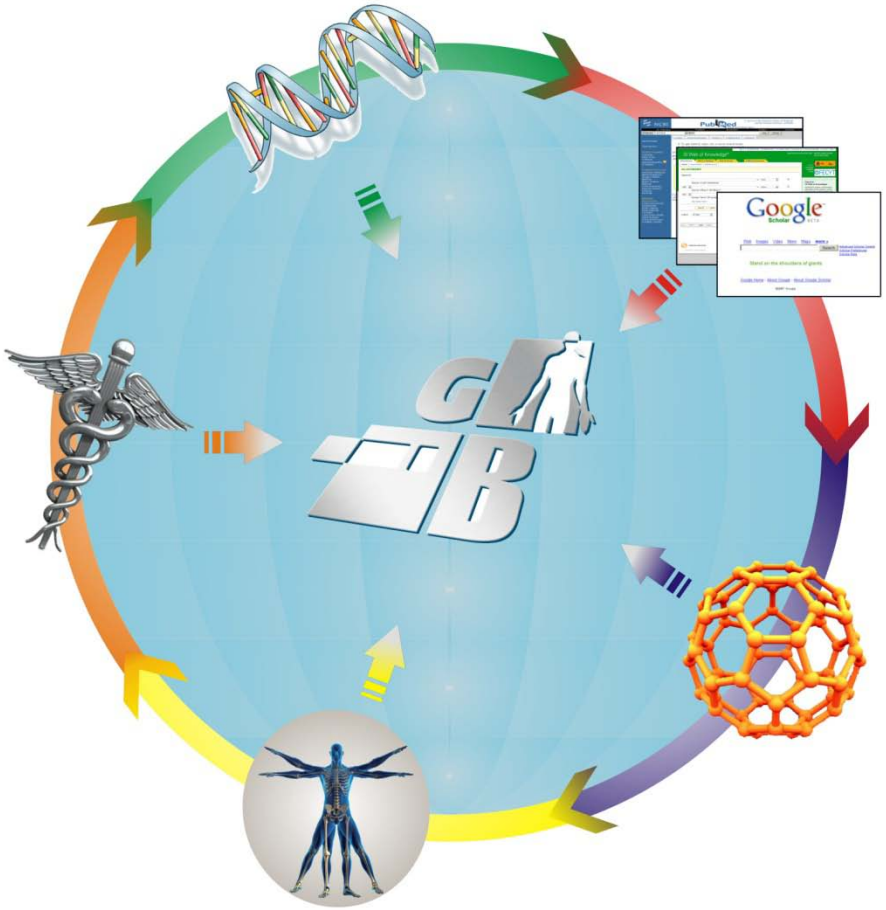




Challenges for Medical Informatics /Bioinformatics interactions (Maojo and Kulikowski, JAMIA, Nov. 2003)

Central ideas: 1st, Biomedicine, around the concept of "information"

"-omics"



In multiple heterogeneous sources

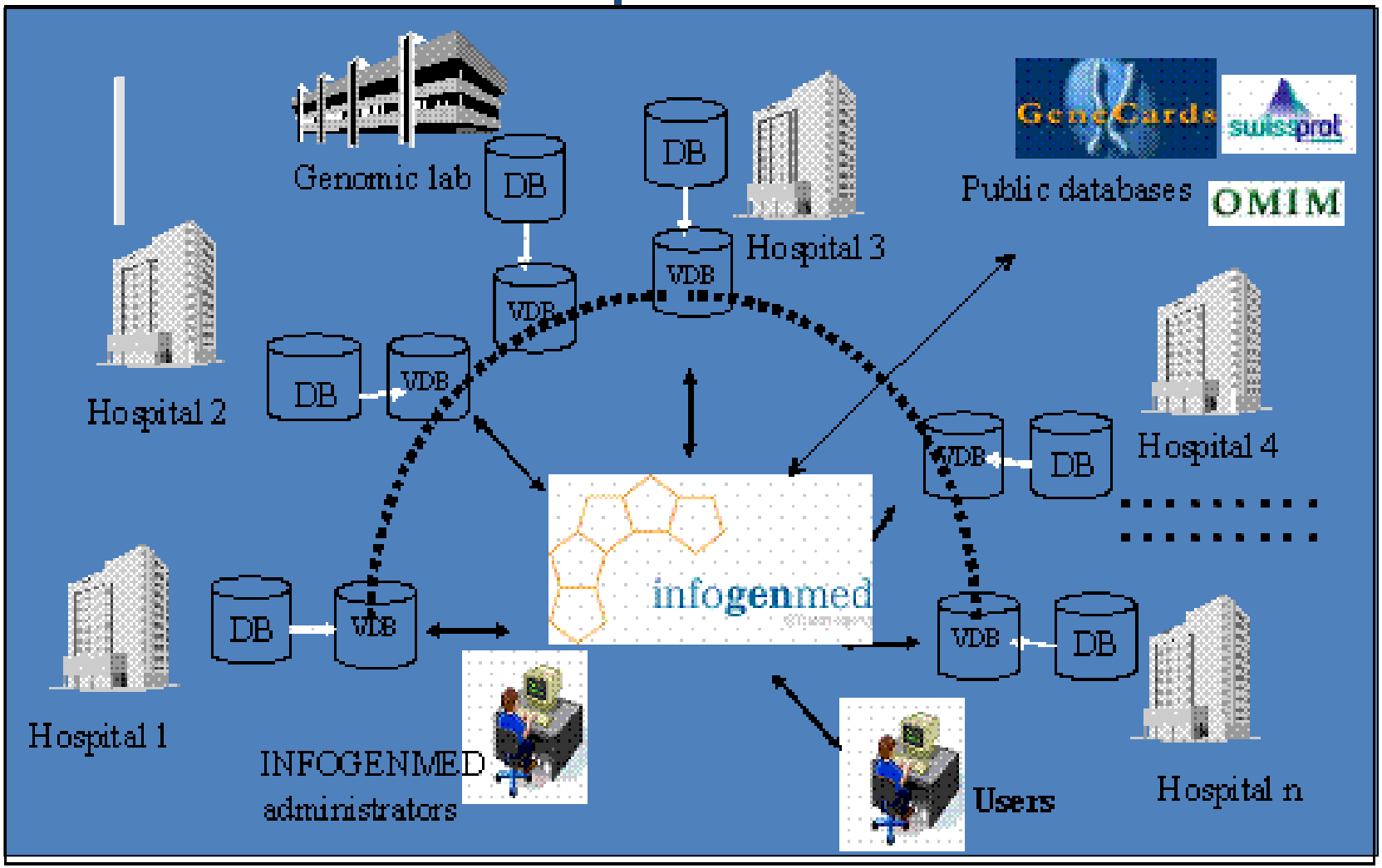
Medical (clinical, public health, etc)

In nanomedicine

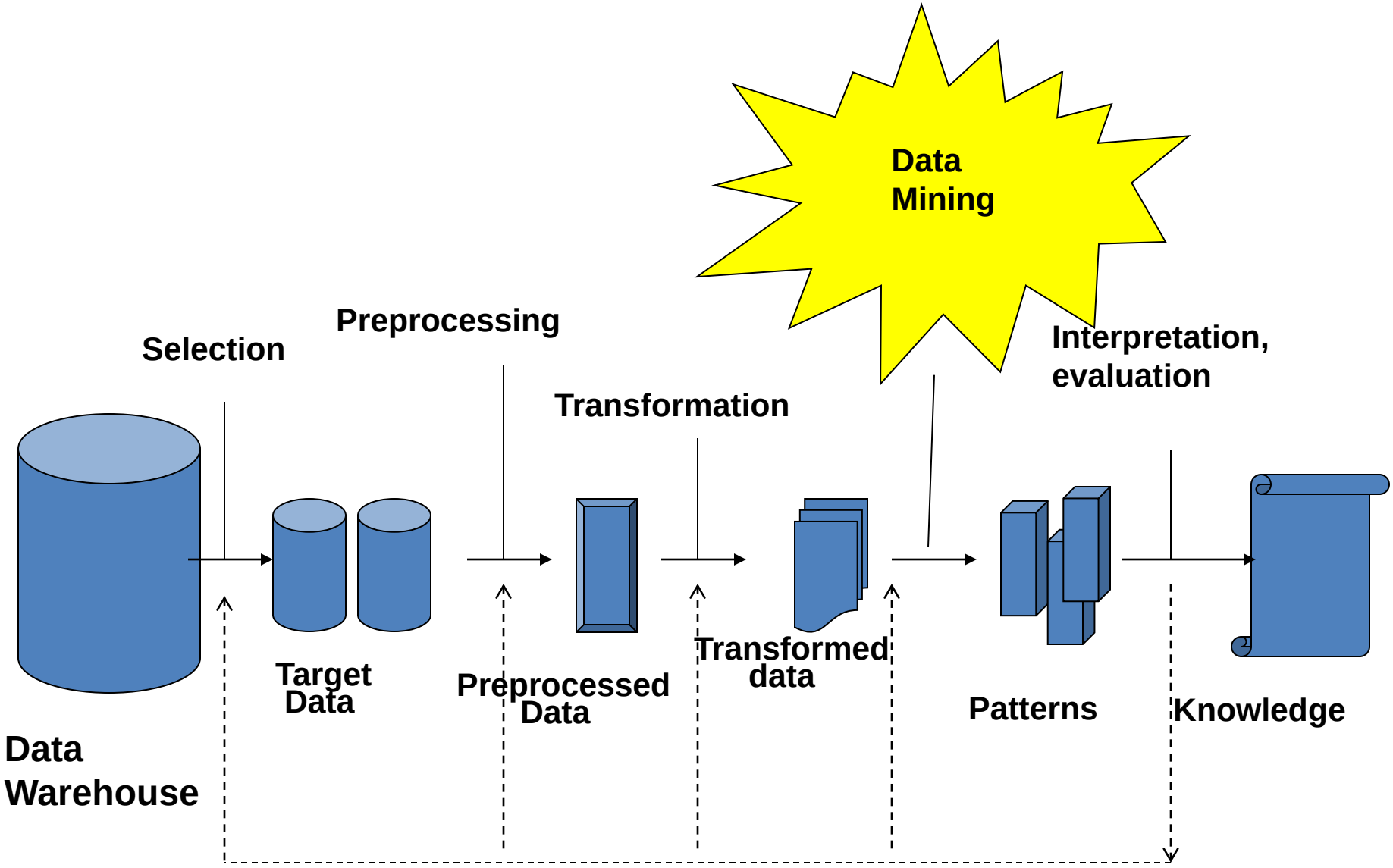
Simulation ("Virtual Physiological Human" program)

Central ideas

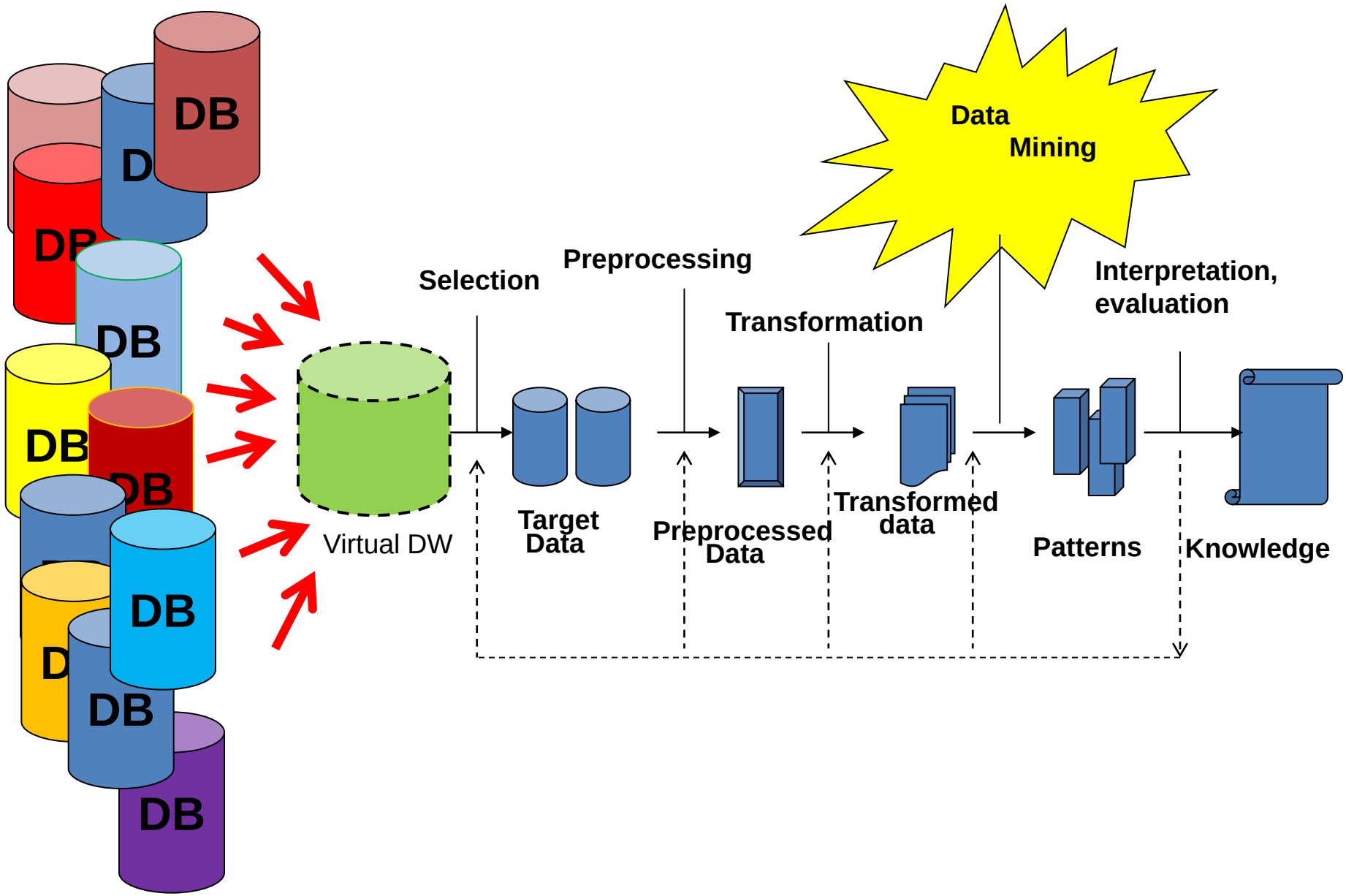
2nd: To facilitate the open access and intelligent use of information in biomedical practice and research



OntoFusion: a pioneering idea of linking medical and –omics information from heterogeneous sources (2002)

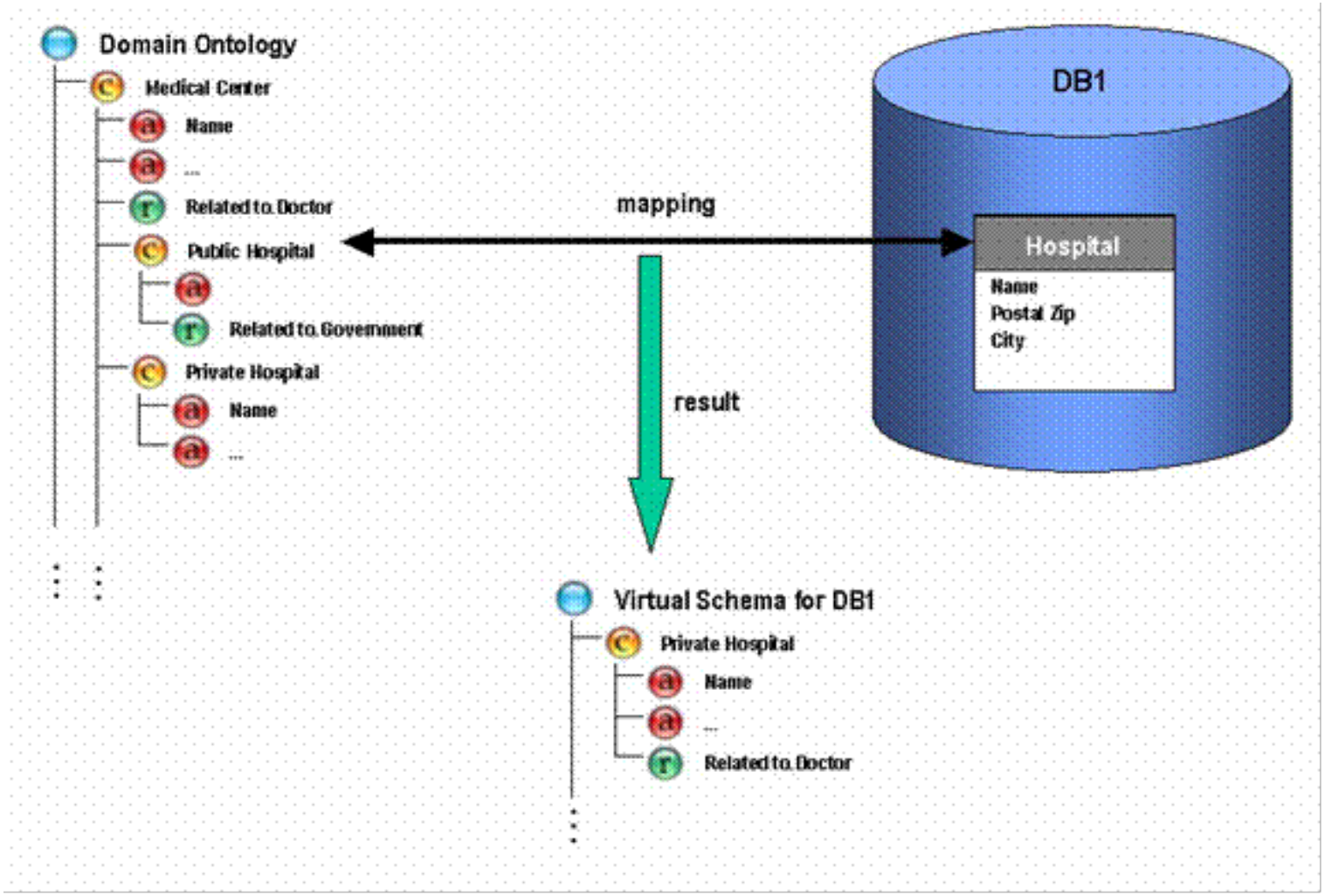


Knowledge Discovery in Databases: classical (data warehouse) methodology

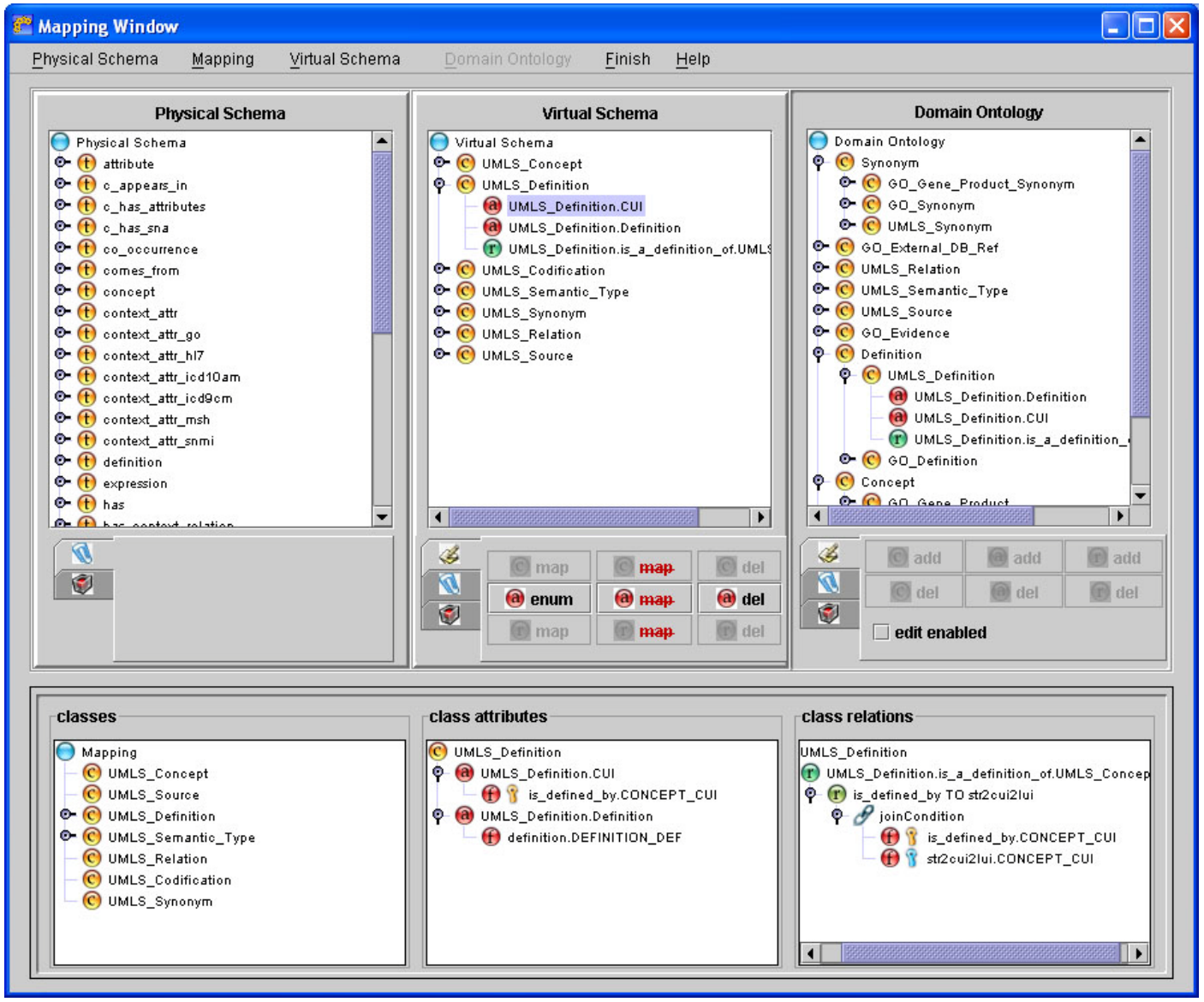


Knowledge Discovery in Databases: distributed approach

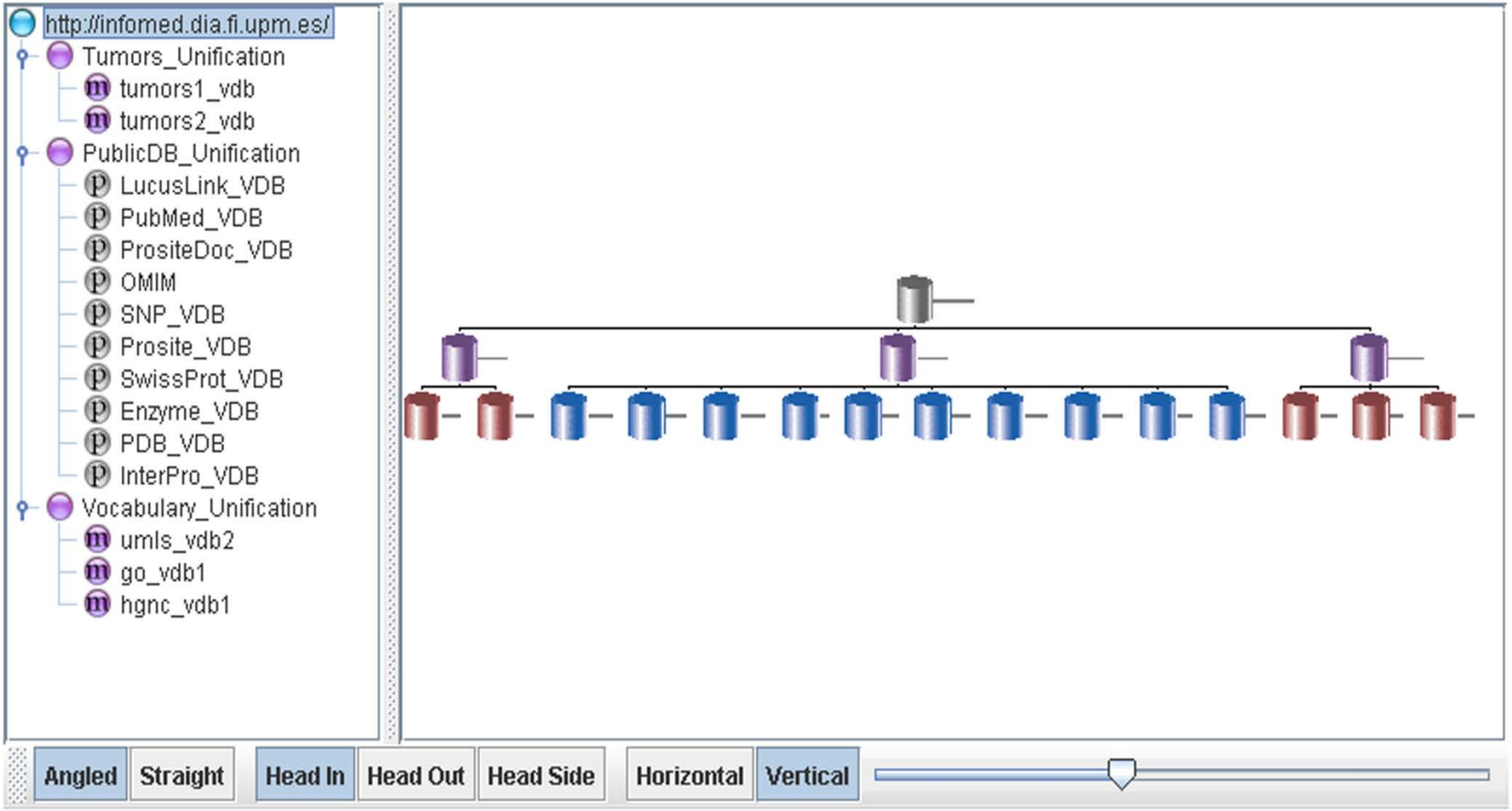
Homogeneization model (using biomedical ontologies)



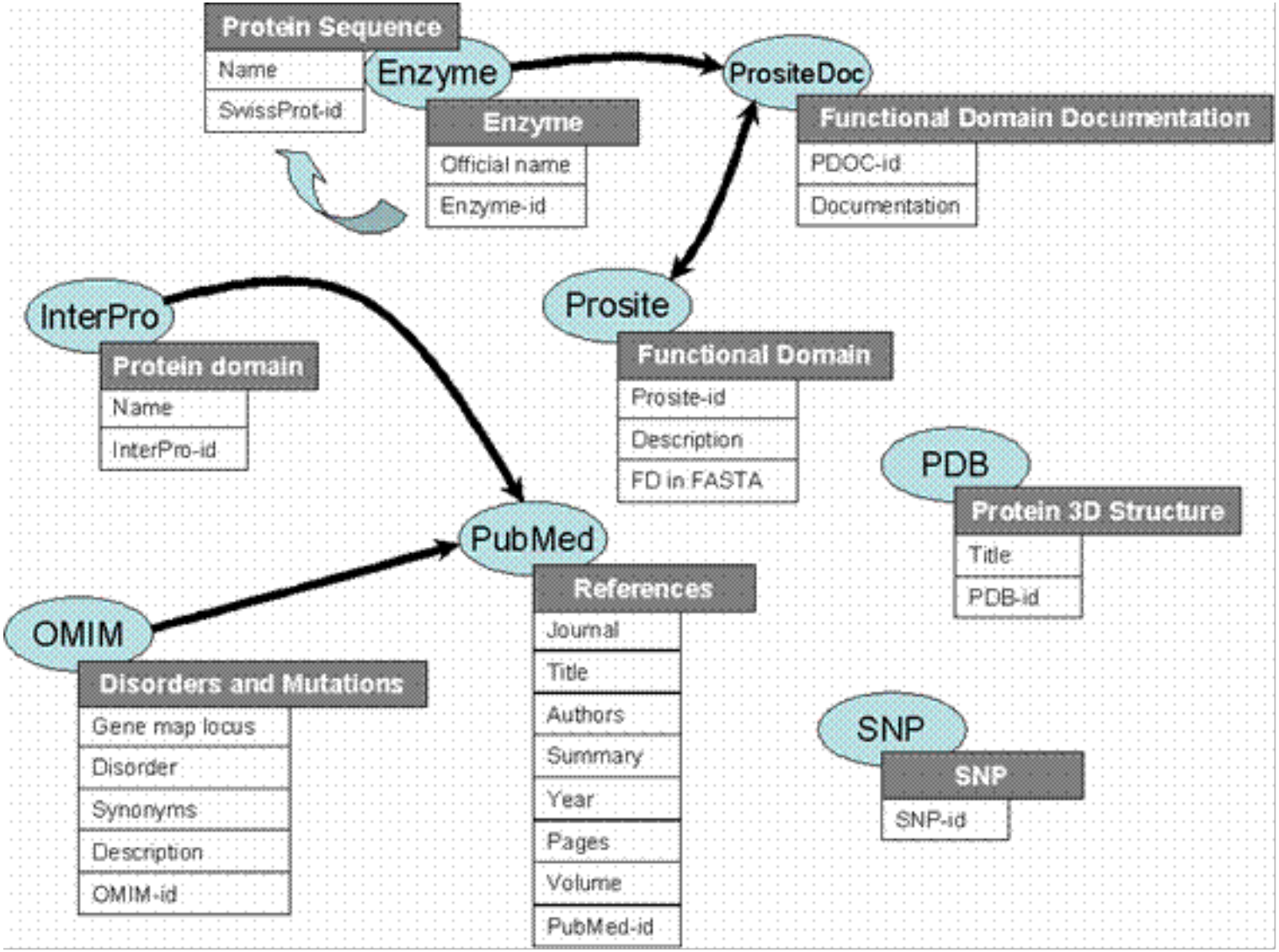
Mapping tools



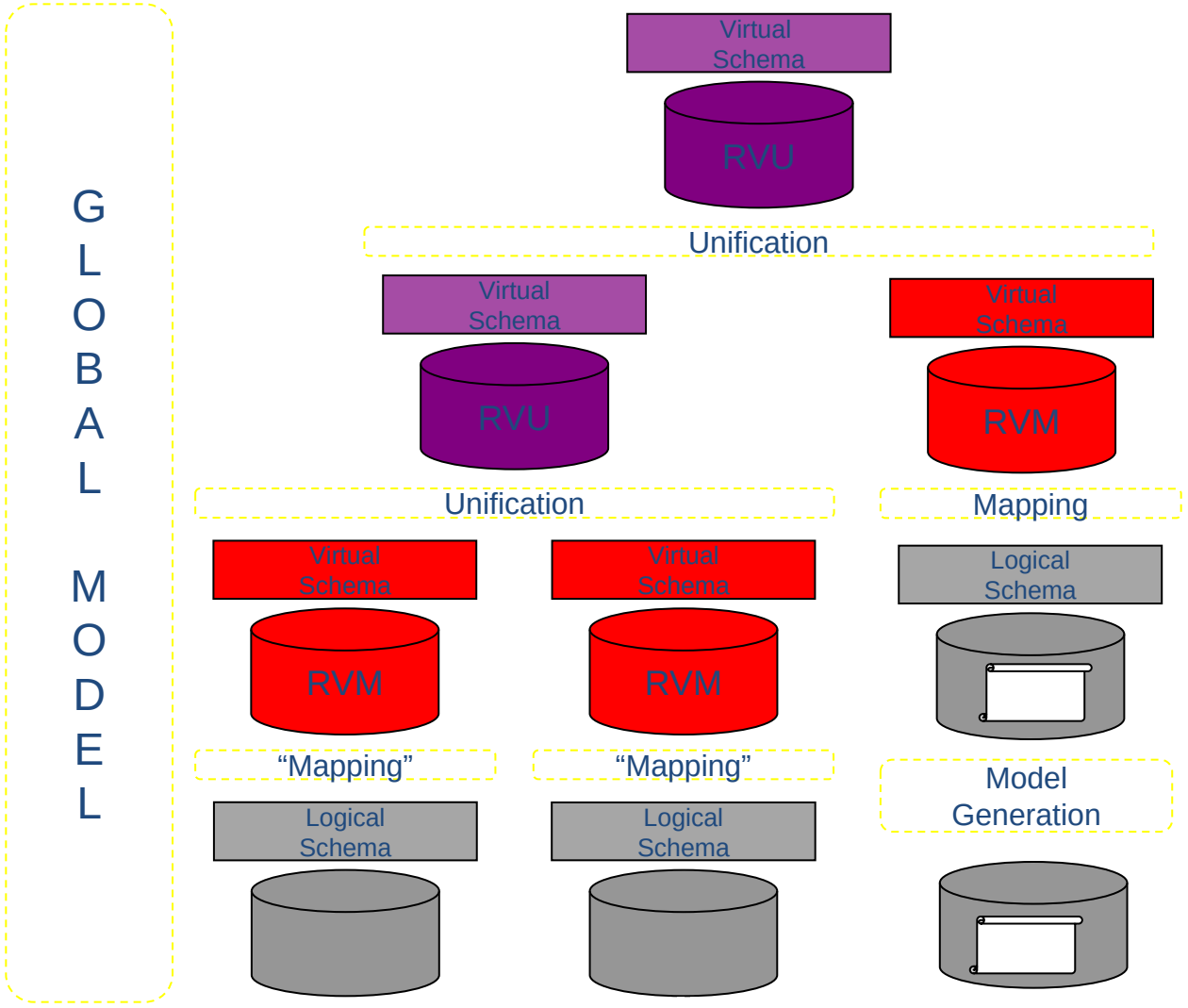
OntoFusion: Clinical and Genomic database integration



Establishing links between public databases



Data + Text mining: Integration of Structured and Non-Structured sources (ONTOFUSION)

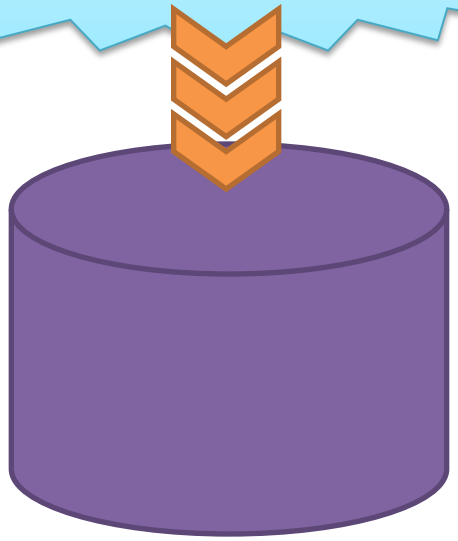


Applications: Integration of Structured and Non-Structured sources (ONTOFUSION)

- Unified Model
  - 257 concepts.
  - 106 hierarchical relationships
  - 425 “ad-hoc” relationships

Subset of the Unified Model

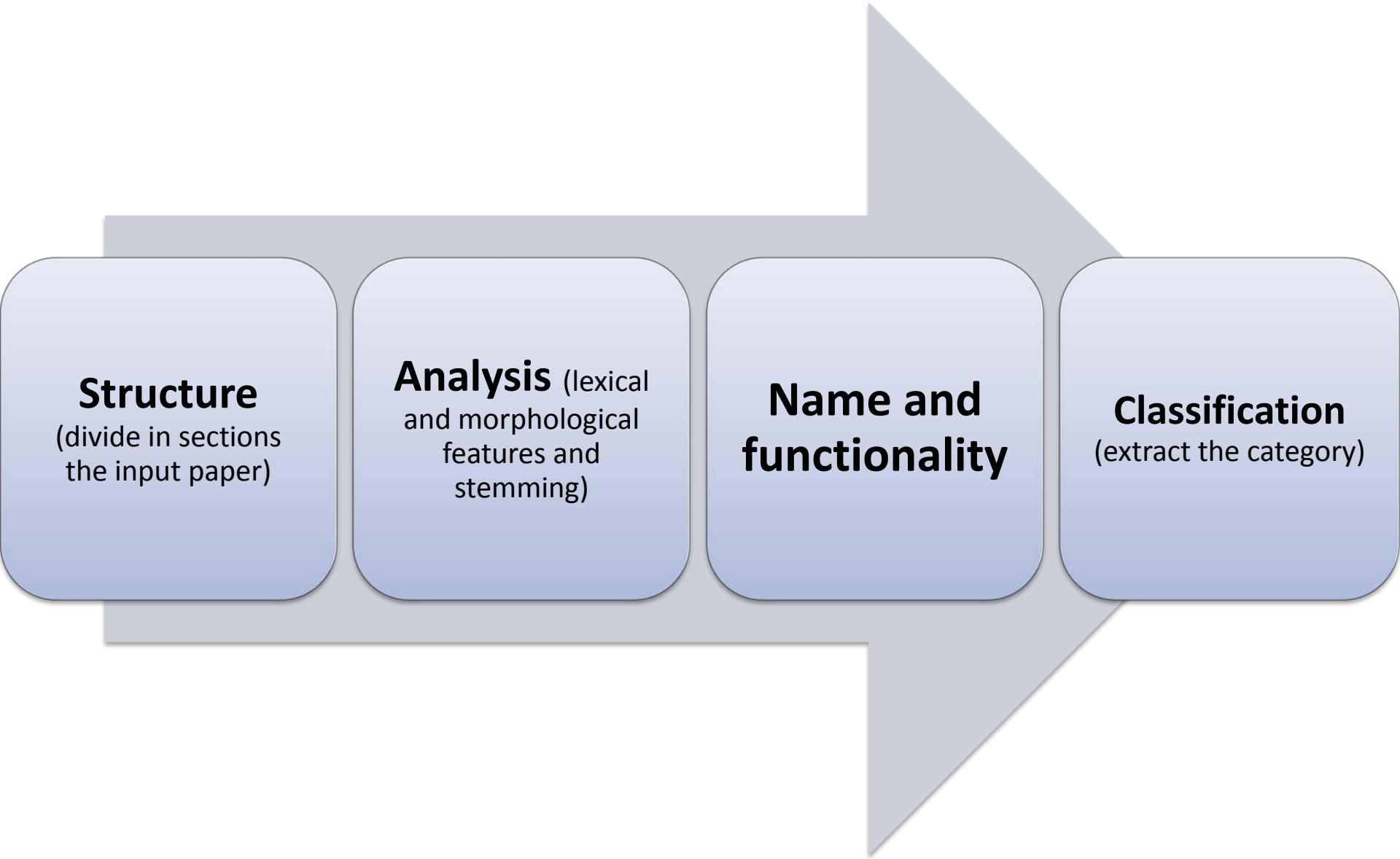
Automatic Extraction of Resources' Information from the literature (like PubMed & PMC)



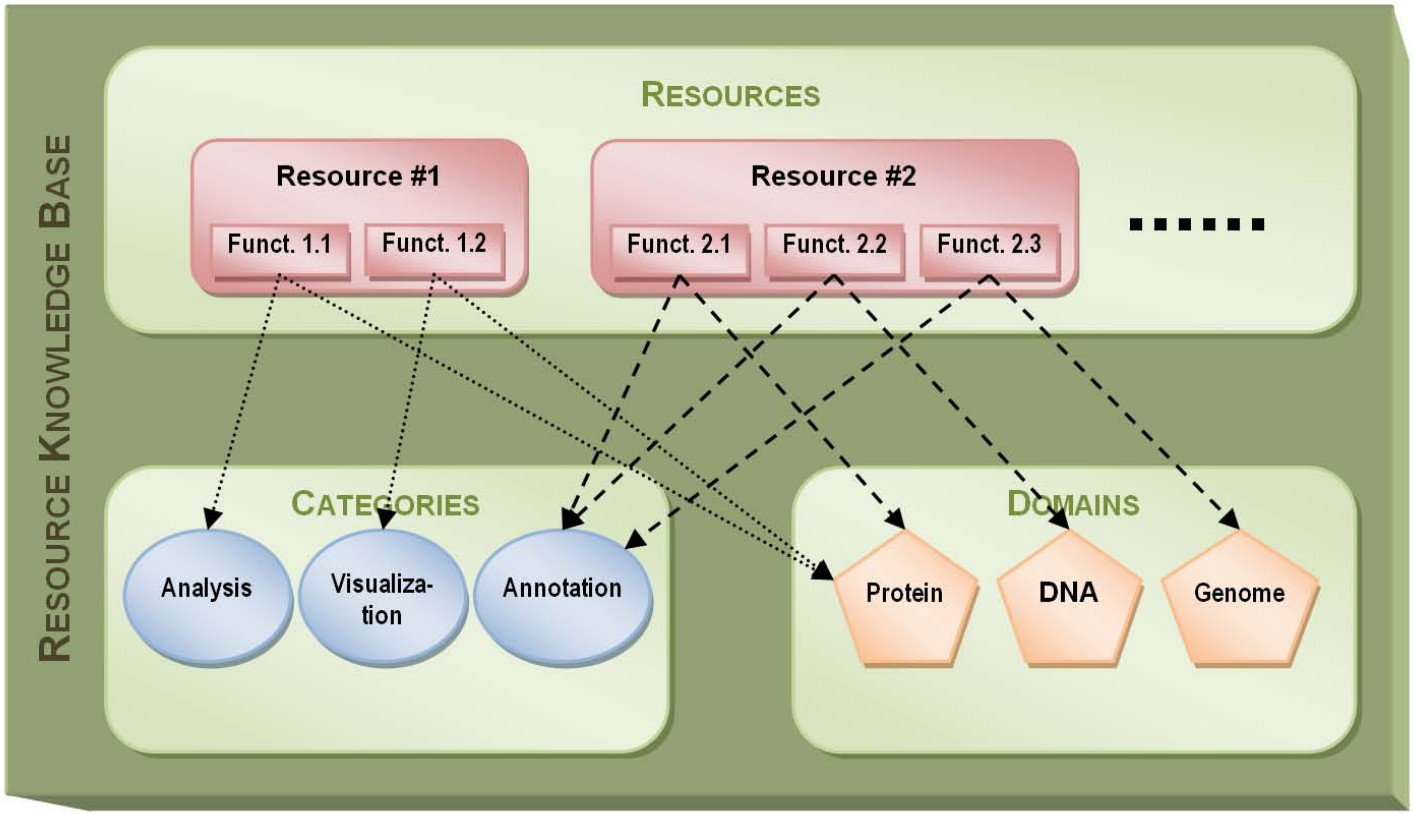
Public Open Source BioMedical Informatics Resources Inventory

Cork, Sept 2011, Victor Maojo

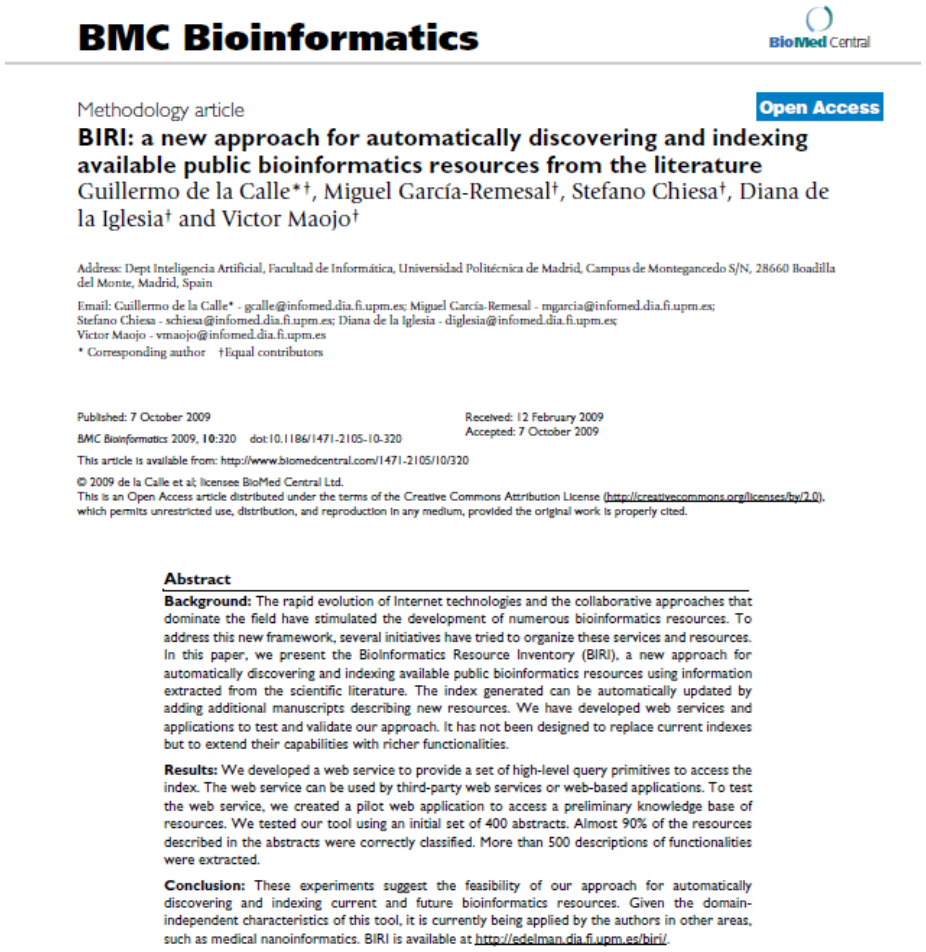
The Analysis process



How to build the Resource Knowledge Base



Towards a bioinformatics resourceome (open resources)



BMC Bioinformatics

BioMed Central

Methodology article

Open Access

BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature

Guillermo de la Calle*†, Miguel García-Remesal†, Stefano Chiesa†, Diana de la Iglesia† and Victor Maojo†

Address: Dept Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo S/N, 28660 Boadilla del Monte, Madrid, Spain

Email: Guillermo de la Calle* - gcalle@infomed.dia.fi.upm.es; Miguel García-Remesal - mgarcia@infomed.dia.fi.upm.es; Stefano Chiesa - schiesa@infomed.dia.fi.upm.es; Diana de la Iglesia - diglesia@infomed.dia.fi.upm.es; Victor Maojo - vmaojo@infomed.dia.fi.upm.es

* Corresponding author †Equal contributors

Published: 7 October 2009

BMC Bioinformatics 2009, **10**:320 doi:10.1186/1471-2105-10-320

Received: 12 February 2009
Accepted: 7 October 2009

This article is available from: http://www.biomedcentral.com/1471-2105/10/320

Abstract

Background: The rapid evolution of Internet technologies and the collaborative approaches that dominate the field have stimulated the development of numerous bioinformatics resources. To address this new framework, several initiatives have tried to organize these services and resources. In this paper, we present the BioInformatics Resource Inventory (BIRI), a new approach for automatically discovering and indexing available public bioinformatics resources using information extracted from the scientific literature. The index generated can be automatically updated by adding additional manuscripts describing new resources. We have developed web services and applications to test and validate our approach. It has not been designed to replace current indexes but to extend their capabilities with richer functionalities.

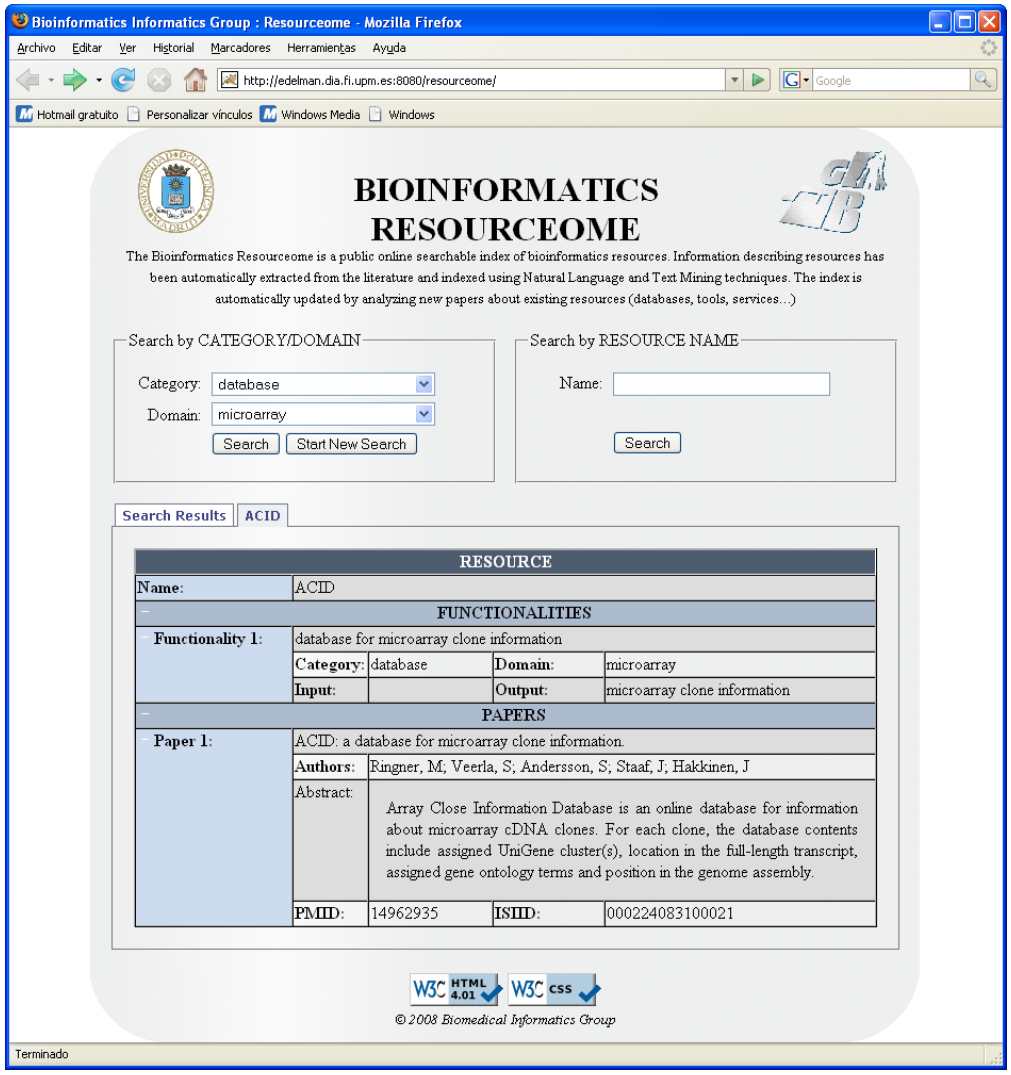
Results: We developed a web service to provide a set of high-level query primitives to access the index. The web service can be used by third-party web services or web-based applications. To test the web service, we created a pilot web application to access a preliminary knowledge base of resources. We tested our tool using an initial set of 400 abstracts. Almost 90% of the resources described in the abstracts were correctly classified. More than 500 descriptions of functionalities were extracted.

Conclusion: These experiments suggest the feasibility of our approach for automatically discovering and indexing current and future bioinformatics resources. Given the domain-independent characteristics of this tool, it is currently being applied by the authors in other areas, such as medical nanoinformatics. BIRI is available at http://edelman.dia.fi.upm.es/biri/.

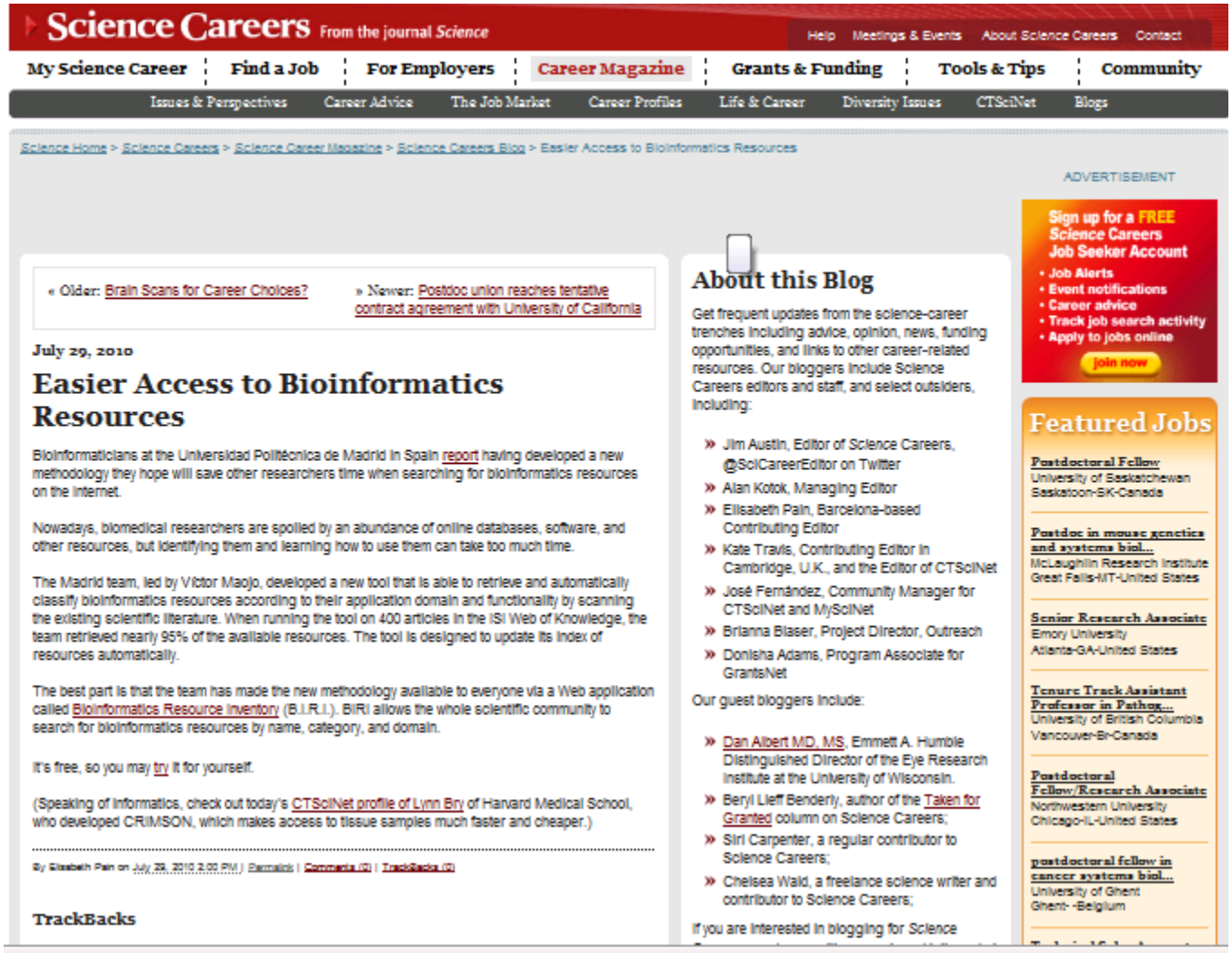
BIRI: A new method for the automatic discovery and indexing of bioinformatics references from the literature, designed to create a repository of resources

The Action-Grid project is funded by the European Commission under the FP7

BIRI: text mining for automatically creating an inventory of bioinformatics resources



Reference in Science



Science Careers From the journal Science

Help Meetings & Events About Science Careers Contact

My Science Career | Find a Job | For Employers | Career Magazine | Grants & Funding | Tools & Tips | Community

Issues & Perspectives | Career Advice | The Job Market | Career Profiles | Life & Career | Diversity Issues | CTSciNet | Blogs

Science Home > Science Careers > Science Career Magazine > Science Careers Blog > Easier Access to Bioinformatics Resources

« Older: Brain Scans for Career Choices?

» Newer: Postdoc union reaches tentative contract agreement with University of California

July 29, 2010

Easier Access to Bioinformatics Resources

Bioinformaticians at the Universidad Politécnica de Madrid in Spain report having developed a new methodology they hope will save other researchers time when searching for bioinformatics resources on the Internet.

Nowadays, biomedical researchers are spoiled by an abundance of online databases, software, and other resources, but identifying them and learning how to use them can take too much time.

The Madrid team, led by Víctor Maojo, developed a new tool that is able to retrieve and automatically classify bioinformatics resources according to their application domain and functionality by scanning the existing scientific literature. When running the tool on 400 articles in the ISI Web of Knowledge, the team retrieved nearly 95% of the available resources. The tool is designed to update its index of resources automatically.

The best part is that the team has made the new methodology available to everyone via a Web application called Bioinformatics Resource Inventory (B.I.R.I.). BIRI allows the whole scientific community to search for bioinformatics resources by name, category, and domain.

It's free, so you may try it for yourself.

(Speaking of informatics, check out today's CTSciNet profile of Lynn Bry of Harvard Medical School, who developed CRIMSON, which makes access to tissue samples much faster and cheaper.)

By Elizabeth Pain on July 29, 2010 2:00 PM | Permalink | Comments (0) | TrackBacks (0)

TrackBacks

About this Blog

Get frequent updates from the science-career trenches including advice, opinion, news, funding opportunities, and links to other career-related resources. Our bloggers include Science Careers editors and staff, and select outsiders, including:

- Jim Austin, Editor of Science Careers, @SciCareerEditor on Twitter
- Alan Kotok, Managing Editor
- Elisabeth Pain, Barcelona-based Contributing Editor
- Kate Travis, Contributing Editor in Cambridge, U.K., and the Editor of CTSciNet
- José Fernández, Community Manager for CTSciNet and MySciNet
- Brianna Blaser, Project Director, Outreach
- Donisha Adams, Program Associate for GrantsNet

Our guest bloggers include:

- Dan Albert MD, MS, Emmett A. Humble Distinguished Director of the Eye Research Institute at the University of Wisconsin.
- Beryl Lieff Benderly, author of the Taken for Granted column on Science Careers;
- Siri Carpenter, a regular contributor to Science Careers;
- Chelsea Wald, a freelance science writer and contributor to Science Careers;

If you are interested in blogging for Science

ADVERTISEMENT

Sign up for a FREE Science Careers Job Seeker Account

- Job Alerts
- Event notifications
- Career advice
- Track job search activity
- Apply to jobs online

join now

Featured Jobs

Postdoctoral Fellow
University of Saskatchewan
Saskatoon-SK-Canada

Postdoc in mouse genetics and systems biol...
McLaughlin Research Institute
Great Falls-MT-United States

Senior Research Associate
Emory University
Atlanta-GA-United States

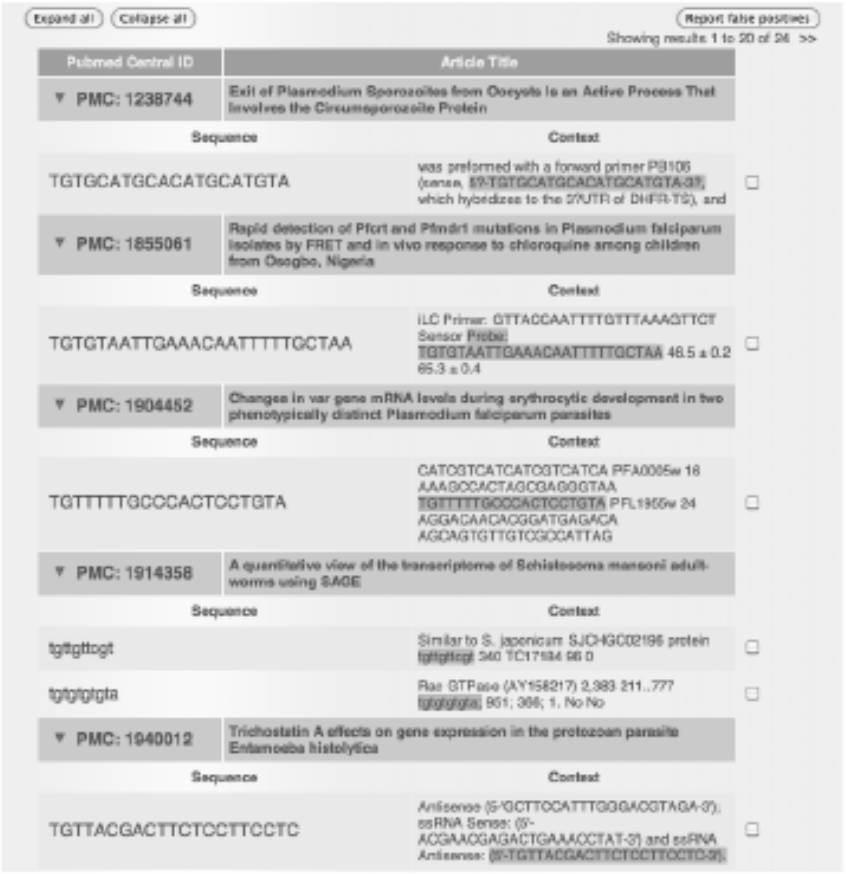
Tenure Track Assistant Professor in Pathog...
University of British Columbia
Vancouver-Br-Canada

Postdoctoral Fellow/Research Associate
Northwestern University
Chicago-IL-United States

postdoctoral fellow in cancer systems biol...
University of Ghent
Ghent- -Belgium

Cork, Sept 2011, Victor Maojo

PubDNA Finder: searching automatically DNA sequences in the literature



BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 21 2010, pages 2801–2802
doi:10.1093/bioinformatics/btq520

Databases and ontologies

Advance Access publication September 9, 2010

PubDNA Finder: a web database linking full-text articles to sequences of nucleic acids

Miguel García-Remesal[1,2,*], Alejandro Cuevas[2], David Pérez-Rey[1,2], Luis Martín[2], Alberto Anguita[2], Diana de la Iglesia[2], Guillermo de la Calle[2], José Crespo[2,3] and Víctor Maojo[1,2]

[1]Departamento de Inteligencia Artificial, Facultad de Informática, [2]Biomedical Informatics Group, Facultad de Informática and [3]Departamento de Lenguajes, Sistemas Informáticos e Ingeniería del Software, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegacedo S/N, 28660 Boadilla del Monte, Madrid, Spain

Associate Editor: John Quackenbush

ABSTRACT

Summary: PubDNA Finder is an online repository that we have created to link PubMed Central manuscripts to the sequences of nucleic acids appearing in them. It extends the search capabilities provided by PubMed Central by enabling researchers to perform advanced searches involving sequences of nucleic acids. This includes, among other features (i) searching for papers mentioning one or more specific sequences of nucleic acids and (ii) retrieving the genetic sequences appearing in different articles. These additional query capabilities are provided by a searchable index that we created by using the full text of the 176 672 papers available at PubMed Central at the time of writing and the sequences of nucleic acids appearing in them. To automatically extract the genetic sequences occurring in each paper, we used an original method we have developed. The database is updated monthly by automatically connecting to the PubMed Central FTP site to retrieve and index new manuscripts. Users can query the database via the web interface provided.

Availability: PubDNA Finder can be freely accessed at http://servet.dia.fi.upm.es:8080/pubdnafinder

Contact: mgarcia@infomed.dia.fi.upm.es

Received on July 16, 2010; revised on August 27, 2010; accepted on September 4, 2010

1 INTRODUCTION

The biological literature is the main source of information reporting empirically validated genetic sequences, such as for instance PCR primers and probes. As result, researchers usually need to review the available literature to search for sequence data, which can be a hard and time-consuming task. PubMed Central is currently the main source of open-access full-text papers reporting genetic sequence data. However, the search engine provided by PubMed Central does not support researchers to retrieve papers containing the genetic sequences specified by the user, and to automatically identify and extract the sequences of nucleic acids mentioned in the retrieved articles.

PubDNA Finder is an online repository linking PubMed Central manuscripts to the different genetic sequences appearing in them. It extends the search capabilities provided by PubMed Central by allowing researchers to (i) retrieve all articles containing the genetic sequences specified by the user—featuring both exact and approximate matching; (ii) retrieve all the sequences appearing in the manuscripts matching a keyword-based query; and (iii) retrieve all articles matching a keyword-based query and containing the sequences specified by the user. PubDNA Finder currently contains the 176 672 papers available from PubMed Central at the time of writing. The database is automatically updated on a monthly basis to retrieve and index new manuscripts.

2 METHODS

To create the index, we downloaded all the 176 672 XML-formatted manuscripts available from the PubMed Central FTP site[1] at the time of writing. We used Apache Lucene[2] 3.0.1 to index the different documents based on the full text of the manuscripts and the genetic sequences appearing in each manuscript. The latter were automatically identified and extracted—together with the context in which they appeared—using a method created by the authors and reported elsewhere (García-Remesal *et al.*, 2010). The adopted method resorts to a rule-based expert system to automatically identify and extract the sequences of nucleic acids. To enable users to interactively query the developed index, we created a web interface.

3 FEATURES

Users can perform three different types of queries using PubDNA Finder, as described below.

3.1 Sequence-based queries

Sequence-based queries (SBQs) are aimed at retrieving all manuscripts containing one or more genetic sequences specified by the user. There are two different types of SBQs: simple and advanced. Simple SBQs are composed of one or more complete sequences linked by a single logical operator, such as 'retrieve all manuscripts containing either the sequence TATGGAAMAGATC-GGCGG or the sequence ATTGGCGAAGTCGGTAGG'.To launch this query, we would type the target sequences—one per line—in the text-box labeled with 'Sequences' (Fig. 1) and then we would select the OR logical operator in the 'Operator' combo box.

*To whom correspondence should be addressed.

[1]ftp://ftp.ncbi.nlm.nih.gov/pub/pmc
[2]http://lucene.apache.org/

García-Remesal et al. BMC Bioinformatics 2010, 11:410
http://www.biomedcentral.com/1471-2105/11/410

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

A method for automatically extracting infectious disease-related primers and probes from the literature

Miguel García-Remesal[1,2*†], Alejandro Cuevas[2†], Victoria López-Alonso[3], Guillermo López-Campos[3], Guillermo de la Calle[2], Diana de la Iglesia[2], David Pérez-Rey[1,2], José Crespo[2,4], Fernando Martín-Sánchez[3], Víctor Maojo[1,2]

Abstract

Background: Primer and probe sequences are the main components of nucleic acid-based detection systems. Biologists use primers and probes for different tasks, some related to the diagnosis and prescription of infectious diseases. The biological literature is the main information source for empirically validated primer and probe sequences. Therefore, it is becoming increasingly important for researchers to navigate this important information. In this paper, we present a four-phase method for extracting and annotating primer/probe sequences from the literature. These phases are: (i) convert each document into a tree of paper sections, (2) detect the candidate sequences using a set of finite state machine-based recognizers, (3) refine problem sequences using a rule-based expert system, and (4) annotate the extracted sequences with their related organism/gene information.

Results: We tested our approach using a test set composed of 297 manuscripts. The extracted sequences and their organism/gene annotations were manually evaluated by a panel of molecular biologists. The results of the evaluation show that our approach is suitable for automatically extracting DNA sequences, achieving precision/recall rates of 97.98% and 95.77%, respectively. In addition, 76.66% of the detected sequences were correctly annotated with their organism name. The system also provided correct gene-related information for 46.18% of the sequences assigned a correct organism name.

Conclusions: We believe that the proposed method can facilitate routine tasks for biomedical researchers using molecular methods to diagnose and prescribe different infectious diseases. In addition, the proposed method can be expanded to detect and extract other biological sequences from the literature. The extracted information can also be used to readily update available primer/probe databases or to create new databases from scratch.

Background

Molecular technologies are used in routine clinical practice to identify microorganisms, and evaluate the presence of virulence factors, antibiotic resistance determinants and host-microbe interactions [1]. For instance, numerous nucleic acid assays have been developed [2] using hybridization or DNA extension techniques that include a wide range of technologies, such as polymerase chain reaction (PCR) methods [3], gene and whole genome sequencing [4,5], Luminex [6] and microarray analysis [7].

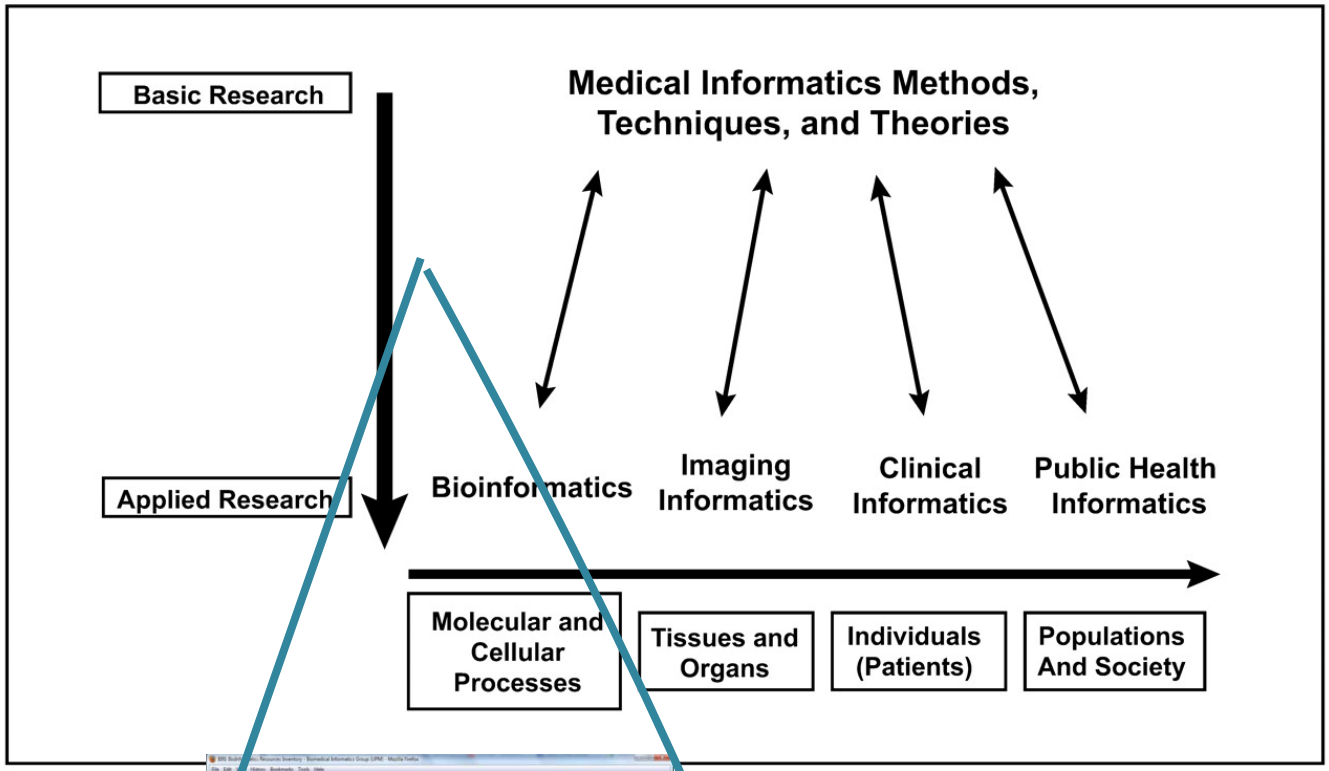
There is a wide range of technologies that provide specific short base sequences of DNA as probes — used to detect the complementary base sequence of interest— or as primers—that guide the DNA amplification process—used for different purposes. Primers and probes are the main components of nucleic acid-based detection systems and have been the subject of multiple studies. Therefore, different software programs have been developed to design these specific sequences of primers and probes minimizing potential cross-hybridization to be spotted, for example, as oligonucleotides in cDNA

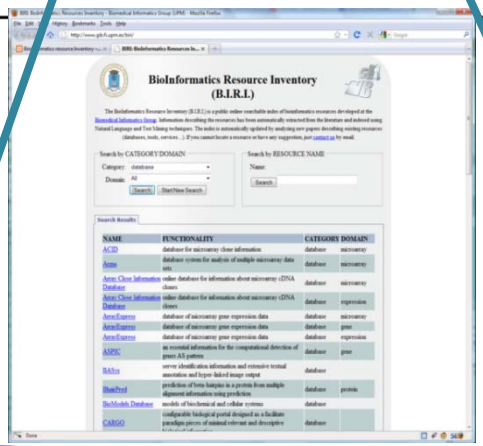
* Correspondence: mgarcia@infomed.dia.fi.upm.es
† Contributed equally
Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegacedo S/N, 28660 Boadilla del Monte, Madrid, Spain
Full list of author information is available at the end of the article

Expanding the work on text mining to other Bioinformatics topics

Cork, Sept 2011, Victor Maojo

Work at the UPM: Building (automatically) inventories of resources including **bioinformatics**…

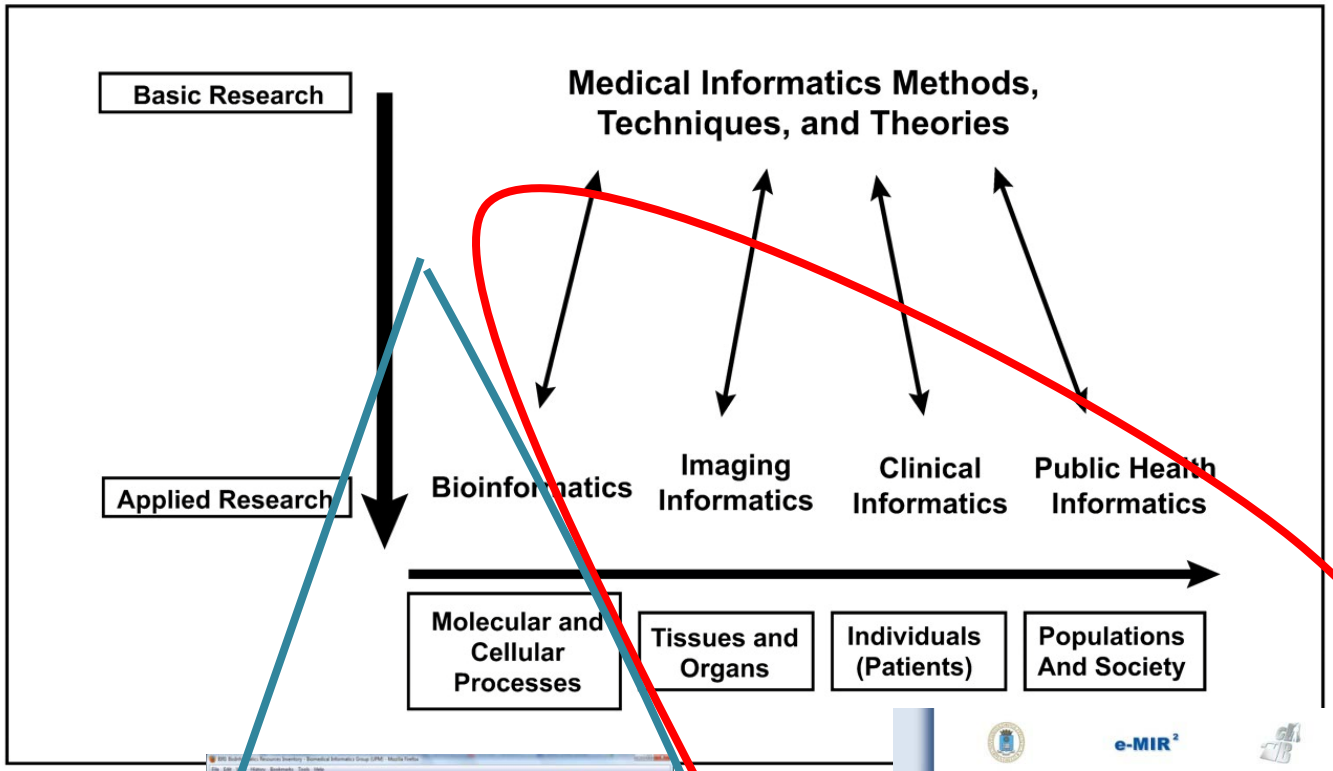




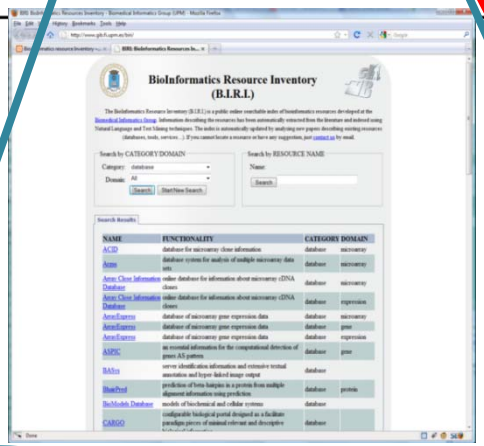
© Ted Shortliffe

Cork, Sept 2011, Victor Maojo

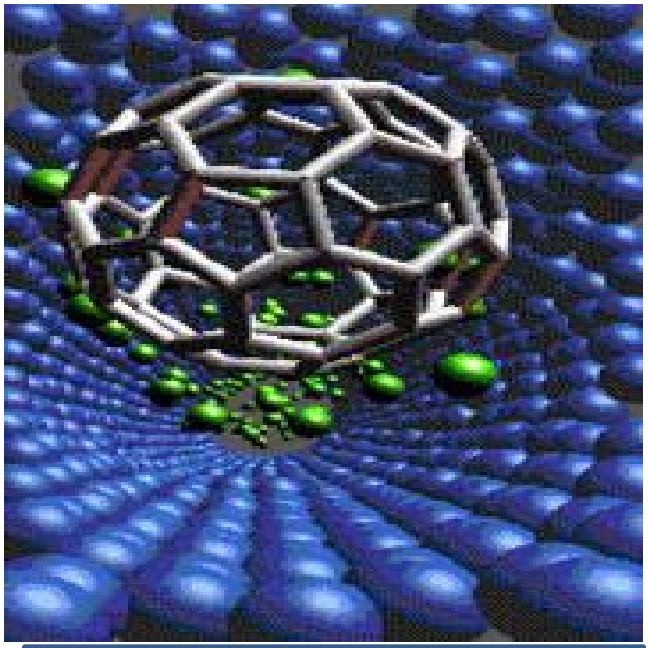
Building (automatically) inventories of resources including bioinformatics, medical…



© Ted Shortliffe



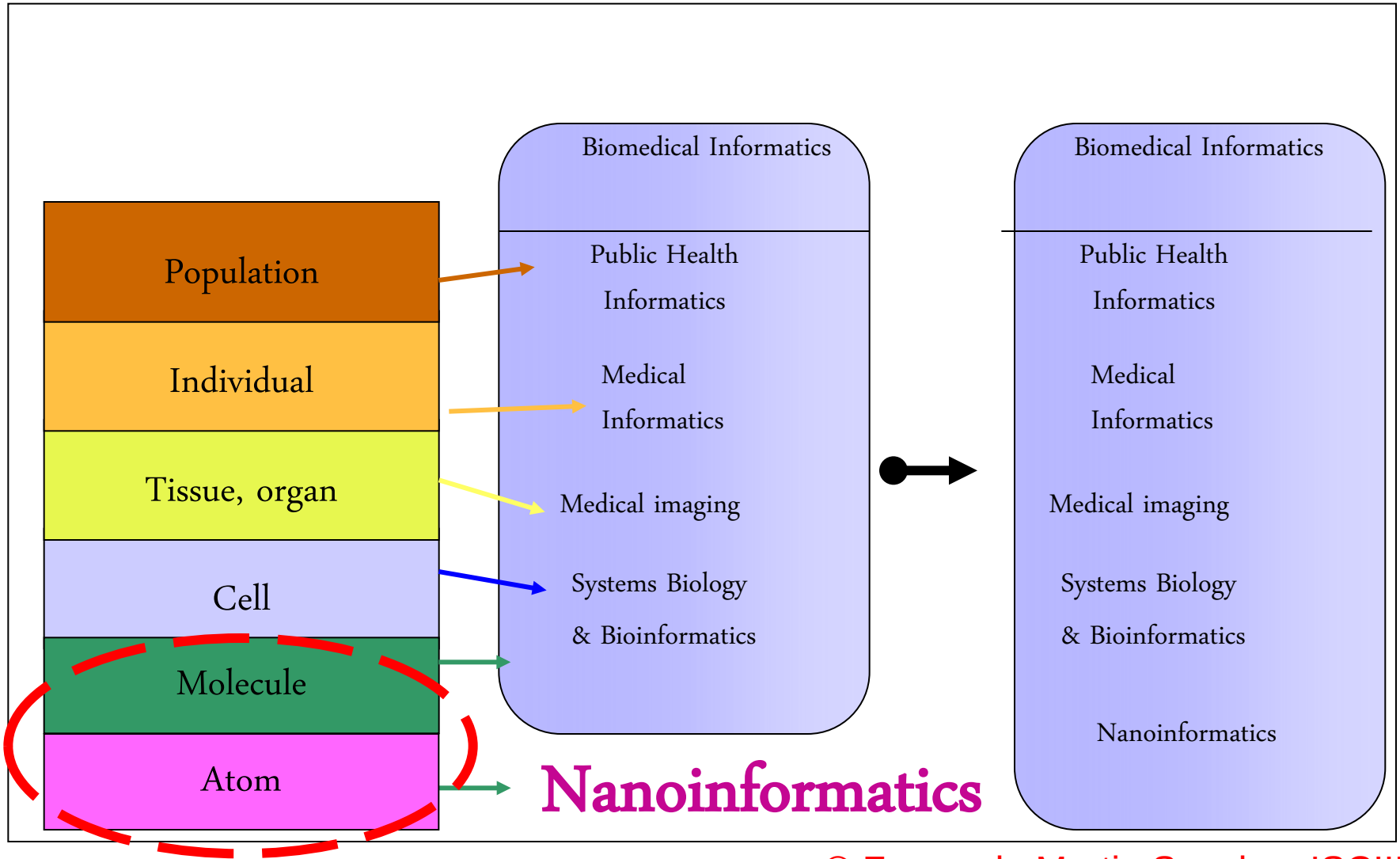
Cork, Sept 2011,
Victor Maojo



Going nano

Extending Biomedical Informatics : Nanoinformatics

(SCOPE) From anatomy (macro) towards microlevels (systems biology) and Nanomedicine



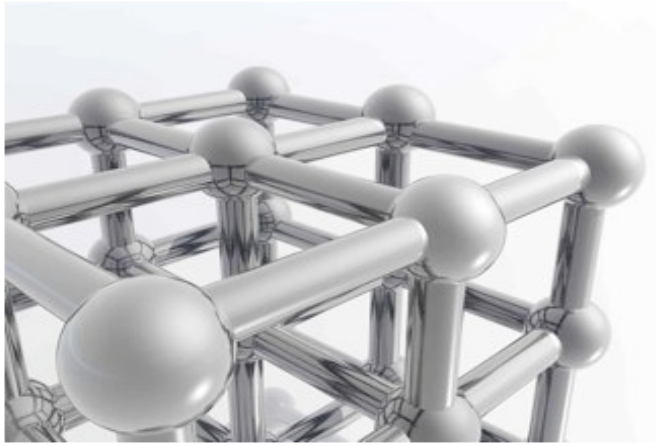
Nanoinformatics: beginning (2007)

Workshop on Nanoinformatics Strategies
June 12-13, 2007, Westin Gateway Hotel, Arlington Virginia
Hosted by the National Nanomanufacturing Network

Agenda (talks & links)

Workshop Purpose

Participants

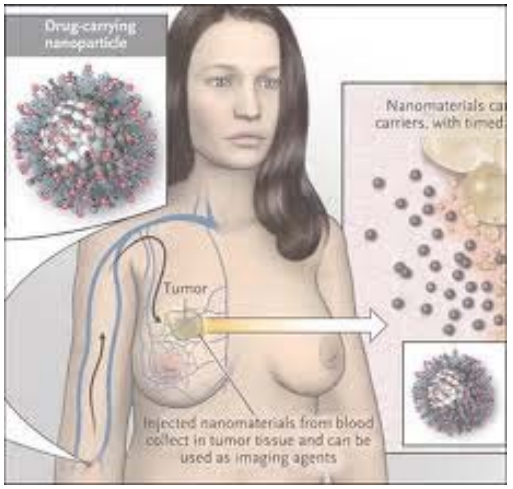


National Nanomanufacturing Network

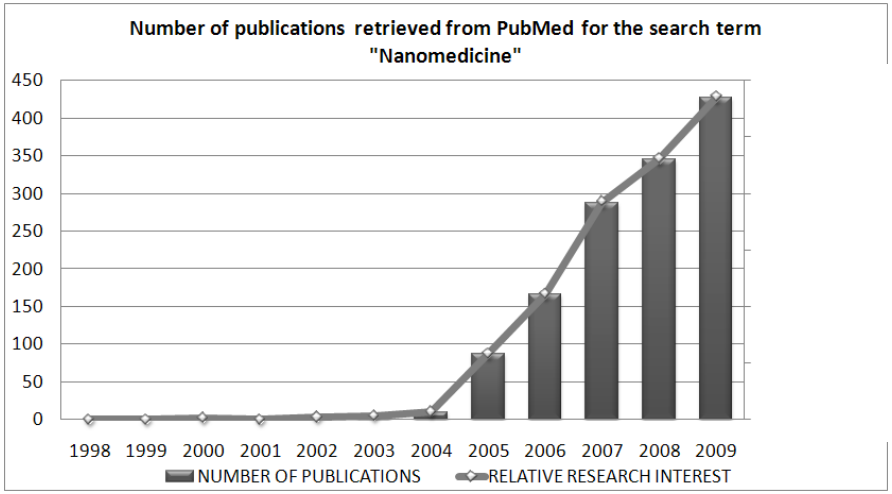
The Workshop on Nanoinformatics Strategies was supported by the National Science Foundation through a grant to the NSF Center for Hierarchical Manufacturing at the University of Massachusetts Amherst.

Nanomedicine:
using nanomaterials for medical care and research

EXAMPLES OF APPLICATIONS OF NANOTECHNOLOGY IN MEDICINE		
Prevention and Diagnosis	Smart sensors	Monitoring and diagnosing disease, controlling intelligent devices, in hospitals or in the home (i.e nanowire systems for diagnosis)
	New methods for molecular imaging	Early detection and diagnosis of disease
	Contrasts	To enhance quality of MRI and other imaging techniques
	Detection of molecules	In vitro diagnostics. Gold nanoparticles for detecting genetic mutations
Therapy	Implantable materials and devices	Tissue repair, and replacement as therapies in regenerative medicine
	Nanorobots	Diagnostic aids, with therapeutic potential
	Drug delivery	Nanoparticles to reach specific targets in the body or for improving solubility characteristics of existing drugs
	Selective therapies	Nanoparticles properties can destroy/control cancerous tissues, for instance by heating
	Gene delivery	Transporting DNA into cells for gene therapy
	Genetic screening	Used in high-throughput detection devices for detecting drug sensitivity
	Inhibiting agents	Delaying the spreading of sexually transmitted diseases (HIV, only in animals yet)



Use of nanoparticles for drug delivery
(© The New England Journal of Medicine, 2010)



Pediatric Research

Wolters Kluwer Health | Lippincott Williams & Wilkins

Pediatric Research:
May 2010 - Volume 67 - Issue 5 - pp 481-489
doi: 10.1203/PDR.0b013e3181d6245e
Nanopediatrics Review Articles: Improved Computer Technology

Nanoinformatics and DNA-Based Computing: Catalyzing Nanomedicine

MAOJO, VICTOR; MARTIN-SANCHEZ, FERNANDO; KULIKOWSKI, CASIMIR; RODRIGUEZ-PATON, ALFONSO; FRITTS, MARTIN

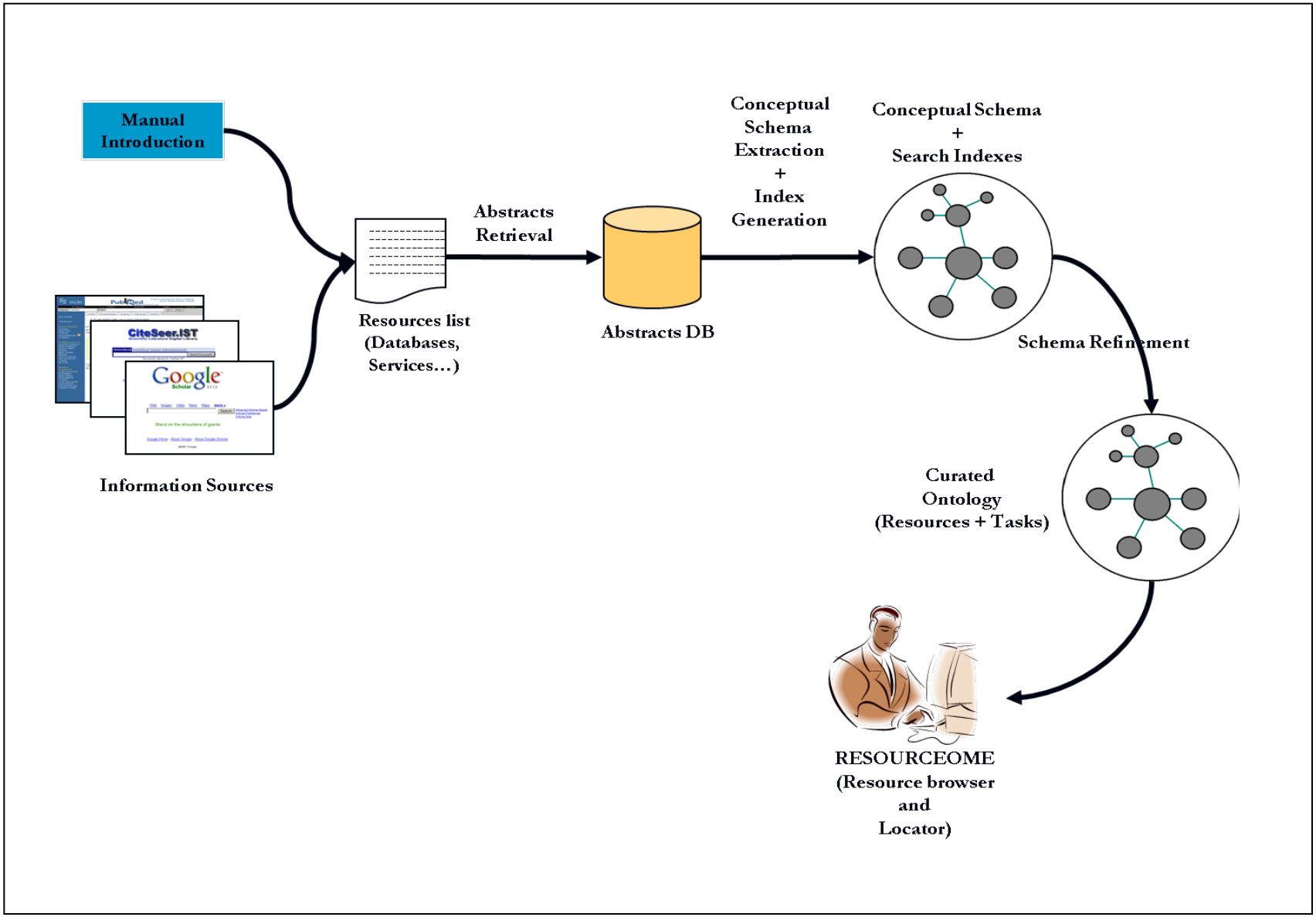
FREE

Article Outline

Author Information

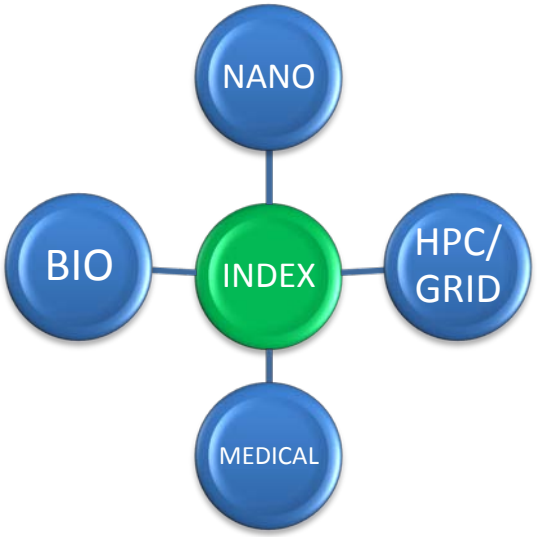
Departamento de Inteligencia Artificial [V.M.], Universidad Politecnica de Madrid, Madrid 28660 Spain; Medical Bioinformatics Department [F.M.-S.], National Institute of Health "Carlos III," Madrid 28220, Spain; Department of Computer Science [C.K.], Rutgers, State University of New Jersey

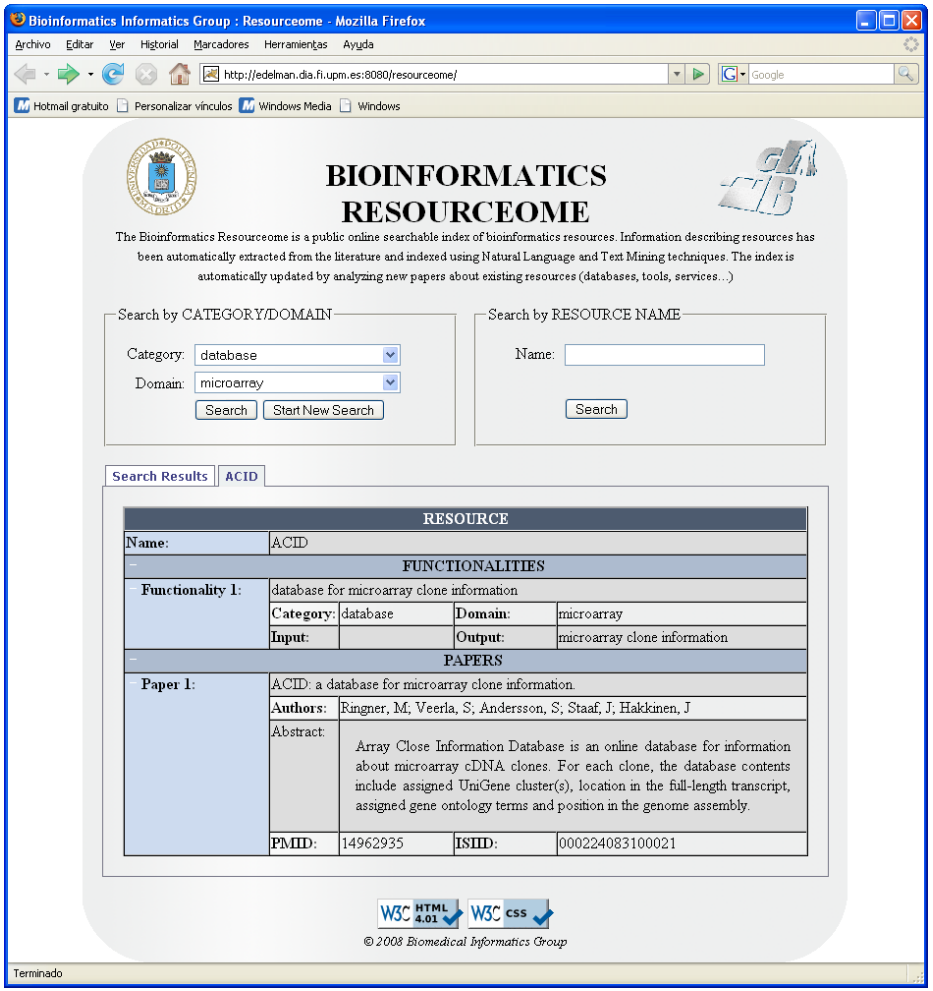
Nano-Resourceome: A repository of Tools and Resources



UPM: Building an index of Grid/Nano/Bio/ Medical resources

- Huge number of resources publicly available over the Internet (databases, services and tools).
- Compiles annotated resources and their associated information
- Compiles annotated resources and their associate information





The Action-Grid project is funded by the European Commission under the FP7

Problems of current biomedical ontologies (our own perspective)

- For us, actually, many other problems arise, (see the debate in the Methods of Information in Medicine journal (June 2011)

- In this case, we believe that current qualitative approaches, based on semantic conceptualizations (and philosophical assunptions in the Open Biomedical Ontologies Foundry) cannot address many problems related to visual and spatial information in biomedicine (e.g., shapes and structures of viruses, molecules or nanoparticles) which are fundamental for living beings


Original Articles © Schattauer 2011 203

Biomedical Ontologies: Toward Scientific Debate

V. Maojo[a]; J. Crespo[a]; M. García-Remesal[a]; D. de la Iglesia[a]; D. Pérez-Rey[a]; C. Kulikowski[b]

[a]Biomedical Informatics Group, Universidad Politécnica de Madrid, Madrid, Spain; [b]Department of Computer Science, Rutgers University, New Jersey, USA

Keywords

Biomedical ontologies, biomedical informatics, spatial ontologies, artificial intelligence, mathematical morphology

Summary

Objectives: Biomedical ontologies have been very successful in structuring knowledge for many different applications, receiving widespread praise for their utility and potential. Yet, the role of computational ontologies in scientific research, as opposed to knowledge management applications, has not been extensively discussed. We aim to stimulate further discussion on the advantages and challenges presented by biomedical ontologies from a scientific perspective.

Methods: We review various aspects of biomedical ontologies going beyond their practical successes, and focus on some key scientific questions in two ways. First, we analyze and discuss current approaches to improve biomedical ontologies that are based largely on classical, Aristotelian ontological models of reality. Second, we raise various open questions about biomedical ontologies that require further research, analyzing in more detail those related to visual reasoning and spatial ontologies.

Results: We outline significant scientific issues that biomedical ontologies should consider, beyond current efforts of building practical consensus between them. For spatial ontologies, we suggest an approach for building "morphospatial" taxonomies, as an example that could stimulate research on fundamental open issues for biomedical ontologies.

Conclusions: Analysis of a large number of problems with biomedical ontologies suggests that the field is very much open to alternative interpretations of current work, and in need of scientific debate and discussion that can lead to new ideas and research directions.

Correspondence to: Victor Maojo, Biomedical Informatics Group, Departamento de Inteligencia Artificial, Facultad de Informática

Methods Inf Med 2011; 50: 203–214
doi: 10.3414/ME10-45-0004
received: November 14, 2010
accepted: January 12, 2011
prepublished: March 21, 2011

"modulates the activity of a caspase, any of a group of cysteine proteases involved in apoptosis" [4]. With the above one can define facts (properties of relationships), instances (individuals belonging to a class), formal axioms, rules, functions, procedures, ontology mappings and other means of manipulating the elements of an ontology. In addition, inheritance in computational ontologies allows properties associated with a higher level (more encompassing) class to be inherited by its subclasses. Over the past years, computational ontologies have been implemented using different ontology mark-up schemas and languages with the goal of transitioning the existing WWW into the Semantic Web [1]. These include EDR, RDF Schema, OIL, DAML+OIL, or the Web Ontology Language (OWL) – a "de facto" current standard [5]. In addition, "upper ontologies" are used to describe general concepts that are shared across various knowledge domains, with the idea of supporting semantic interoperability between different ontologies at lower levels. There are several upper ontologies, each one differing greatly in terms of their users, topics, focus and ontological foundations. [illegible]lude the Basic Formal [illegible], the Descriptive Ontol- [illegible]and Cognitive Engineer- [illegible]eneralized Upper Model [illegible]oC 1101, Process Speci-

Original Articles © Schattauer 2011 217

Discussion of "Biomedical Ontologies: Toward Scientific Debate"

M. Brochhausen[a]; A. Bangor[b]; W. Ceusters[c]; A. Hasman[d]; T. T. Leong[e]; M. Musen[f]; J. L. Oliveira[g]; M. Peleg[h]; A. Rector[i]; S. Schulz[j]

[a]Saarland University, Saarbrücken, Germany; [illegible]; [c]State University at Buffalo, Buffalo NY, USA; [d]AMC University of Amsterdam, Amsterdam, The Netherlands; [e]National University of Singapore, Singapore; [f]Stanford University, Stanford, USA; [g]University of Aveiro, Aveiro, Portugal; [h]University of Haifa, Haifa, Israel; [i]University of Manchester, Manchester, United Kingdom; [j]Medical University of Graz, Graz, Austria

With these comments on the paper "Biomedical Ontologies: Toward scientific debate", written by Victor Maojo et al. [1], Methods of Information in Medicine wants to stimulate a discussion on advantages and challenges of biomedical ontologies. An international group of experts have been invited by the editor of Methods to comment on this paper. Each of the invited commentaries forms one section of this paper.

1. Comment by M. Brochhausen[a]

Biomedical Ontologies and the Scientific Method

The paper Biomedical Ontologies: Toward Scientific Debate [1] seeks to show that current work on biomedical ontologies is deviating from good scientific practice by borrowing insights and methods from "philosophy, linguistic and other non-sciences". In what follows I will argue that its authors hereby apply a double standard. More precisely, I will show that, when we apply to the arguments in their own paper the same critical approach that its authors apply to recent research in biomedical ontologies, then these arguments will be shown to collapse. Indeed this paper itself falls so far short of meeting the standards of good scientific practice that it is not actually a scientific paper at all.

There is agreement that there are certain principles of good scientific practice which are the same for any domain of research, including the humanities [2]. These principles include the acknowledgement of the existing body of work, the requirement that an investigation should be coherently conceived and executed, and that there should be no flaws in methodology [3]. I will show how [1] falls short in respect of all these of these principles.

Acknowledgement of the Existing Body of Work

The authors focus their attentions on the Basic Formal Ontology (BFO) [4, 5], an upper level ontology to support integration of scientific research data through the creation of domain-specific BFO extensions. BFO is now being used for this purpose by some 75 projects almost all of them in the life sciences field, including medicine [6].

One key criticism directed by the authors towards BFO and the ontologies derived from BFO by downward population is the fact that it is labeled as "Aristotelian" by its developers. The argument seems to be that, because Aristotle's works on natural phenomena do not meet the criteria of empirical science as we understand them today, and because some of the basic principles of BFO's ontological realism have – above all as the treatment of definitions – Aristotelian roots, it follows that BFO, too, must be rejected. The level of argument here is, I believe, quite astonishing. The author's reasoning is analogous to an argument to the effect that because, say, Newton's works on alchemy do not meet the criteria of empirical science as we understand them today, then the principles of Newtonian mechanics must likewise be rejected.

In writing about BFO and realism Maojo et al. seem to be unaware of the contemporary debate amongst philosophers between the two camps of realists and nominalists as concerns the status of universals and their role in science [7–9]. They also seem to be unaware of the degree to which the BFO's categorization of reality differs from Aristotle's categorization of the world given in his Categories [10]. The creators and proponents of BFO are proponents of realist doctrine that is standardly referred to in the contemporary literature as "Aristotelian" simply to mark the fact that it rests on a view that there are mind-independent categories instantiated by the particulars in reality. To this, the proponents of BFO add a number of views which have nothing to do with Aristotle, for instance a view to the effect that the way to gain knowledge about categories is to test corresponding general hypotheses by following the scientific method in carrying

Correspondence to: See list of authors' addresses at the end of the article

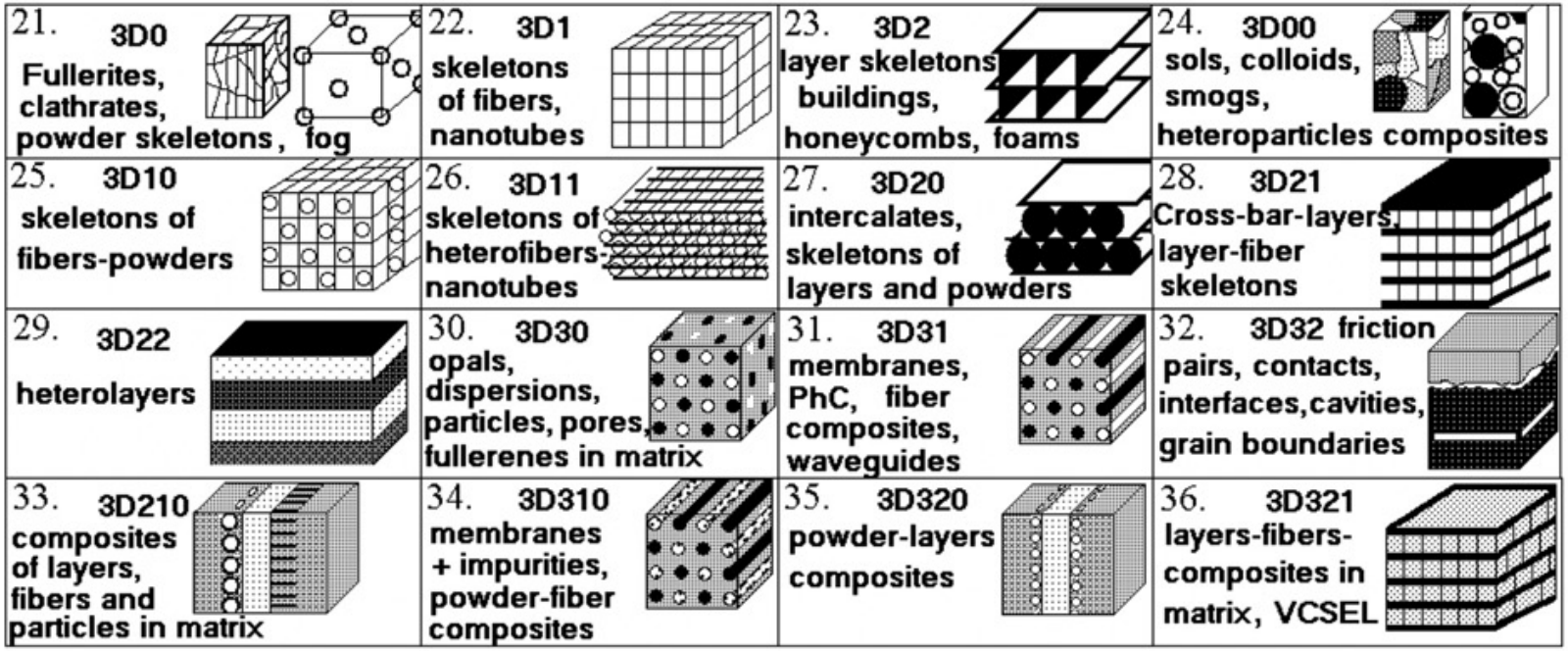
Methods Inf Med 2011; 50: 217–226

1 I want to thank Amanda Hicks for proof-reading and discussing good scientific practice.

Methods Inf Med 3/2011

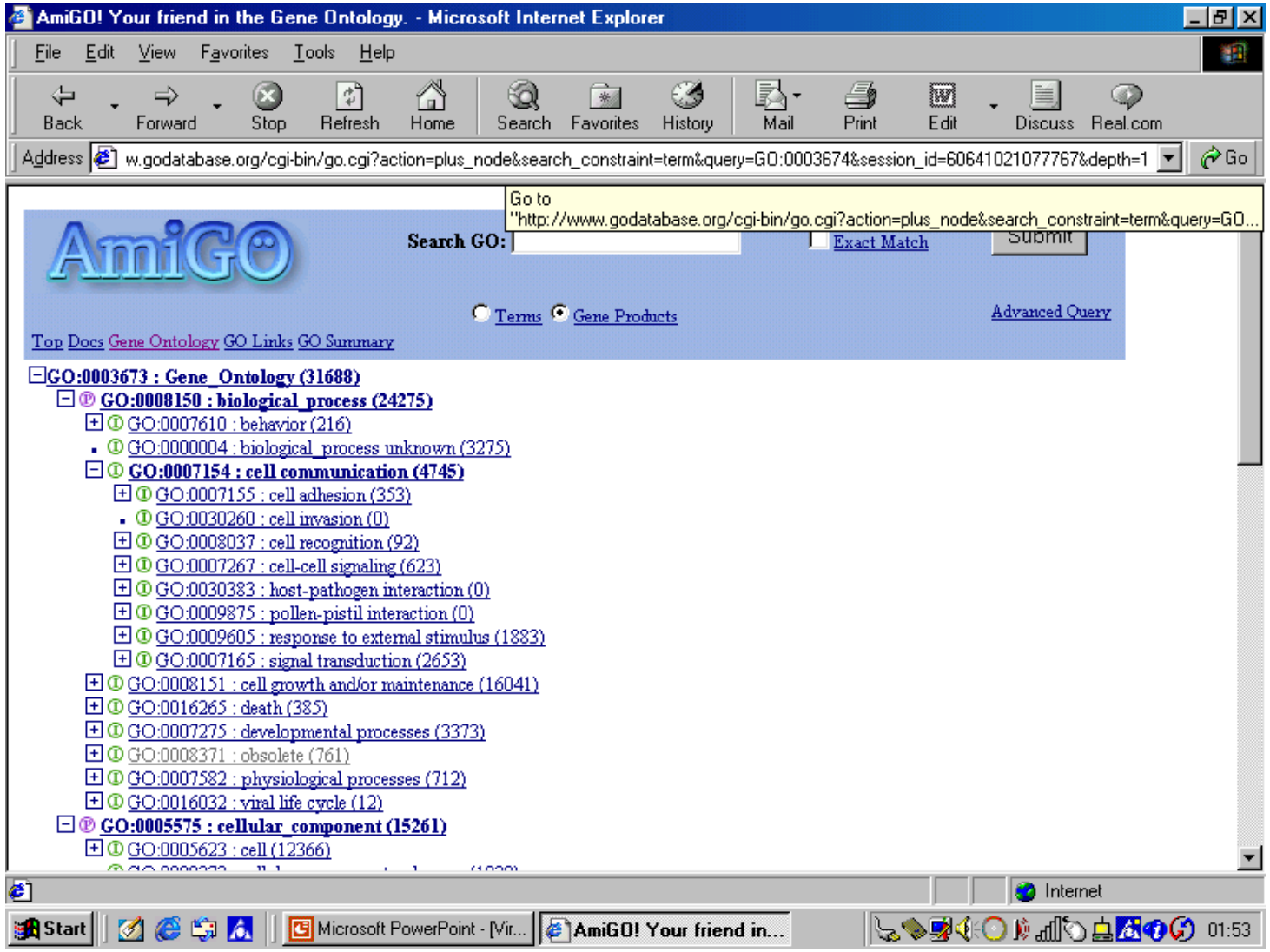
An example
Taxonomies of shapes: Application to nanoparticles

©Pokropivny and Skorokhod



No current, conceptual ontology can capture the complexity of the information of the visual, graphical information embedded in nanoparticles

Gene Ontology http://www.geneontology.org



Light ontology,

mainly useful for annotations and retrieval:

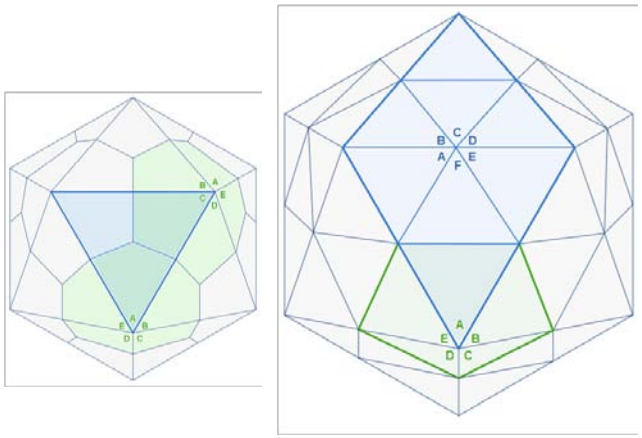
molecular_function
biological_process:
cellular_component

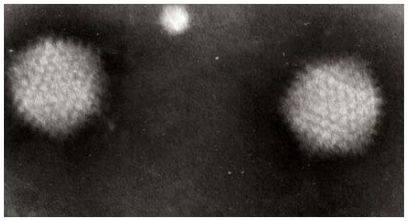
Ontology	Scope	URL	Custodians
Cell Ontology (CL)	cell types from prokaryotes to mammals	obo.sourceforge.net/cgi-bin/detail.cgi?cell	Jonathan Bard, Michael Ashburner, Oliver Hofman
Chemical Entities of Biological Interest (ChEBI)	molecular entities	ebi.ac.uk/chebi	Paula Dematos, Rafael Alcantara
Common Anatomy Reference Ontology (CARO)	anatomical structures in human and model organisms	(under development)	Melissa Haendel, Terry Hayamizu, Cornelius Rosse, David Sutherland,
Foundational Model of Anatomy (FMA)	structure of the human body	fma.biostr.washington.edu	JLV Mejino Jr., Cornelius Rosse
Functional Genomics Investigation Ontology (FuGO)	design, protocol, data instrumentation, and analysis	fugo.sf.net	FuGO Working Group
Gene Ontology (GO)	cellular components, molecular functions, biological processes	www.geneontology.org	Gene Ontology Consortium
Phenotypic Quality Ontology (PaTO)	qualities of anatomical structures	obo.sourceforge.net/cgi-bin/ detail.cgi?attribute_and_value	Michael Ashburner, Suzanna Lewis, Georgios Gkoutos
Protein Ontology (PrO)	protein types and modifications	(under development)	Protein Ontology Consortium
Relation Ontology (RO)	relations	obo.sf.net/relationship	Barry Smith, Chris Mungall
RNA Ontology (RnaO)	three-dimensional RNA structures	(under development)	RNA Ontology Consortium
Sequence Ontology (SO)	properties and features of nucleic sequences	song.sf.net	Karen Eilbeck

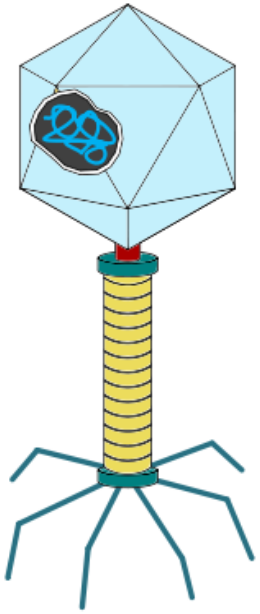
The OBO Foundry

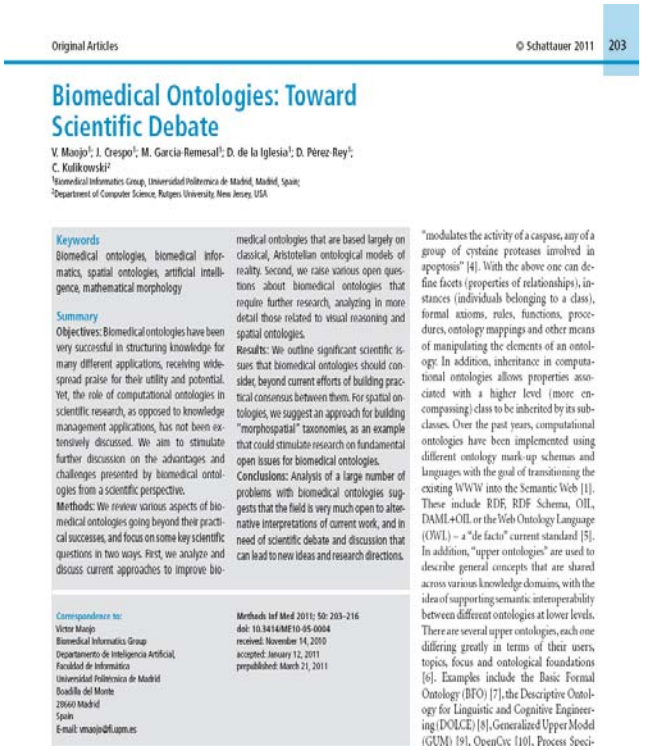
http://obofoundry.org/

Work on pattern classification and nano-ontologies: Taxonomies of shapes and forms









Original Articles © Schattauer 2011 203

Biomedical Ontologies: Toward Scientific Debate

V. Maojo[1]; J. Crespo[1]; M. García-Remesal[1]; D. de la Iglesia[1]; D. Pérez-Rey[1]; C. Kulikowski[2]

[1]Biomedical Informatics Group, Universidad Politécnica de Madrid, Madrid, Spain; [2]Department of Computer Science, Rutgers University, New Jersey, USA

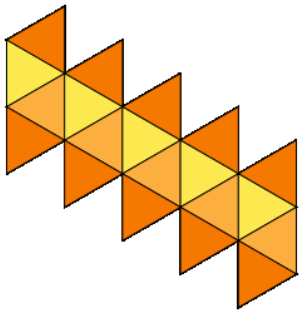
Keywords
Biomedical ontologies, biomedical informatics, spatial ontologies, artificial intelligence, mathematical morphology

Summary
Objectives: Biomedical ontologies have been very successful in structuring knowledge for many different applications, receiving widespread praise for their utility and potential. Yet, the role of computational ontologies in scientific research, as opposed to knowledge management applications, has not been extensively discussed. We aim to stimulate further discussion on the advantages and challenges presented by biomedical ontologies from a scientific perspective.
Methods: We review various aspects of biomedical ontologies going beyond their practical successes, and focus on some key scientific questions in two ways. First, we analyze and discuss current approaches to improve biomedical ontologies that are based largely on classical, Aristotelian ontological models of reality. Second, we raise various open questions about biomedical ontologies that require further research, analyzing in more detail those related to visual reasoning and spatial ontologies.
Results: We outline significant scientific issues that biomedical ontologies should consider, beyond current efforts of building practical consensus between them. For spatial ontologies, we suggest an approach for building "morphospatial" taxonomies, as an example that could stimulate research on fundamental open issues for biomedical ontologies.
Conclusions: Analysis of a large number of problems with biomedical ontologies suggests that the field is very much open to alternative interpretations of current work, and in need of scientific debate and discussion that can lead to new ideas and research directions.

Correspondence to:
Victor Maojo
Biomedical Informatics Group
Departamento de Inteligencia Artificial,
Facultad de Informática
Universidad Politécnica de Madrid
Boadilla del Monte
28660 Madrid
Spain
E-mail: vmaojo@fi.upm.es

Methods Inf Med 2011; 50: 203–216
doi: 10.3414/ME10-05-0004
received: November 14, 2010
accepted: January 12, 2011
prepublished: March 21, 2011

"modulates the activity of a caspase, any of a group of cysteine proteases involved in apoptosis" [4]. With the above one can define facets (properties of relationships), instances (individuals belonging to a class), formal axioms, rules, functions, procedures, ontology mappings and other means of manipulating the elements of an ontology. In addition, inheritance in computational ontologies allows properties associated with a higher level (more encompassing) class to be inherited by its subclasses. Over the past years, computational ontologies have been implemented using different ontology mark-up schemas and languages with the goal of transitioning the existing WWW into the Semantic Web [1]. These include RDF, RDF Schema, OIL, DAML+OIL or the Web Ontology Language (OWL) – a "de facto" current standard [5]. In addition, "upper ontologies" are used to describe general concepts that are shared across various knowledge domains, with the idea of supporting semantic interoperability between different ontologies at lower levels. There are several upper ontologies, each one differing greatly in terms of their users, topics, focus and ontological foundations [6]. Examples include the Basic Formal Ontology (BFO) [7], the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [8], Generalized Upper Model (GUM) [9], OpenCyc [10], Process Speci-



Shapes
•1-D Shapes
•2-D Shapes
— 2-D Geometrical shapes
o2-D Geometrical shapes with genus 0
▪Circles
▪Polygons
•Convex polygons
oSquares
oTriangles
o…
▪Non-convex polygons
o2-D Geometrical shapes with genus 1
—2-D Non-geometrical shapes
•3-D Shapes

With biomedical implications

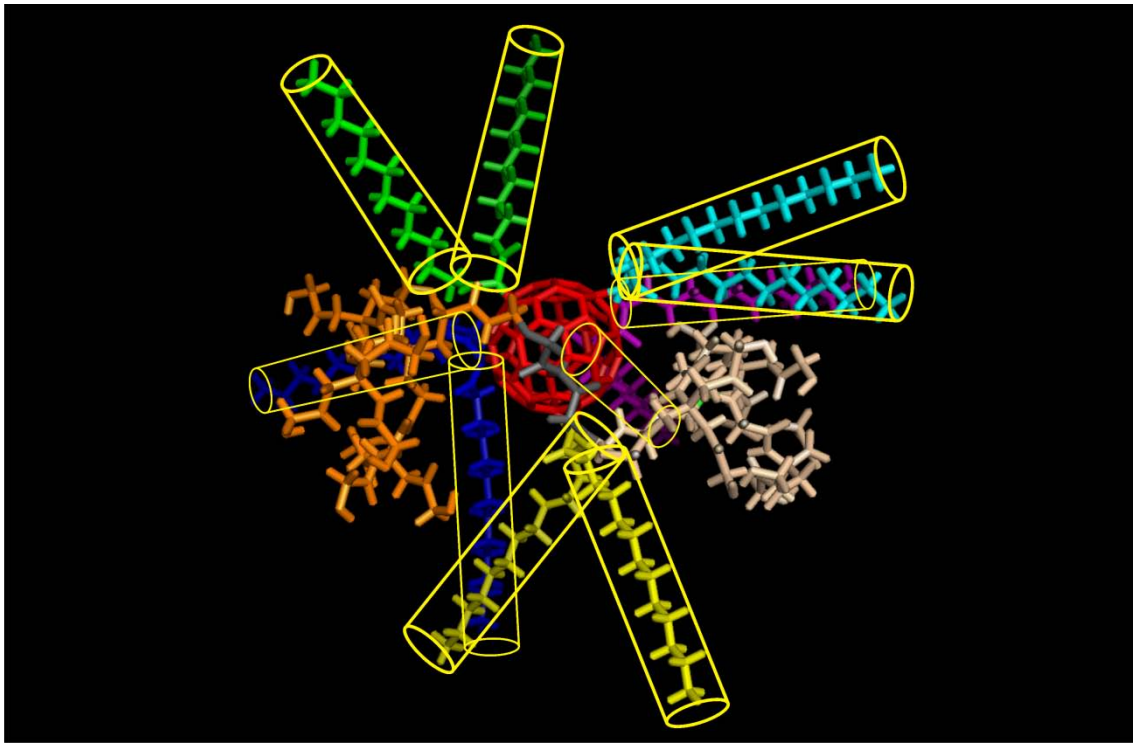


Image processing

Sub-shapes

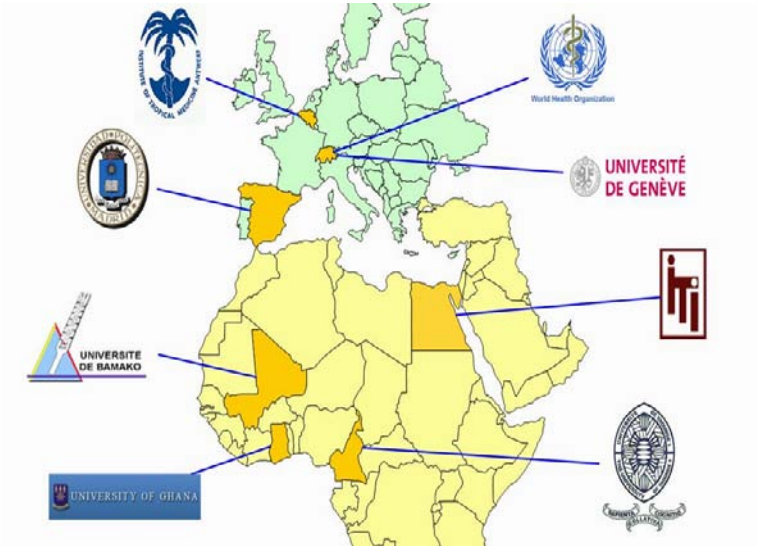
combinations in

major molecules

Image created using
PyMol

10 cylinders / pillars , 1 sphere

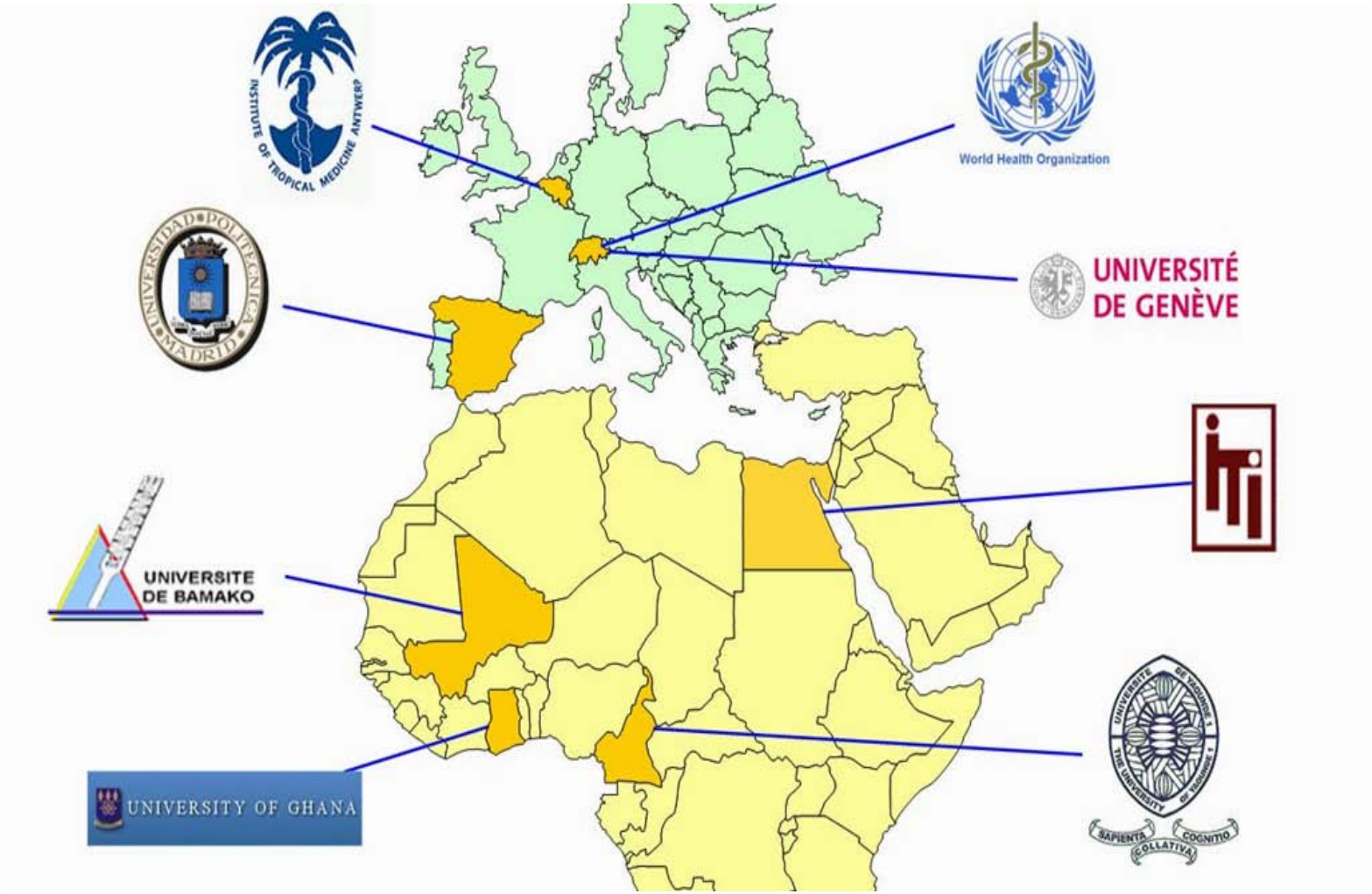
Nanoparticle Source: CSN
NCI Id: 32730



Going South

Using Biomedical Informatics to improve medical research and care in Africa

Going south: the Africa Build scenario



Africa Build: main objectives

- To teach (on-site and through e-learning) African professionals in biomedical informatics and medical topics
- To transfer BMI open resources (databases, tools, services)
- To create virtual communities and Web 2.0 networks of people to work in health issues in Africa — including volunteers
- To build the computing infrastructure needed
- Long term goals: sustainable south-south collaborations

A challenging «proof of concept»: Cloud Computing

Two experiments:

- Madrid-Cairo

- Madrid-Burundi

Alonso-Calvo, R., Crespo, J., Garcia-Remesal, M., Anguita, A. and Maojo, V. On distributing load in cloud computing: A real application for very-large image datasets. Proceedings of ICCS 2010 (in press)

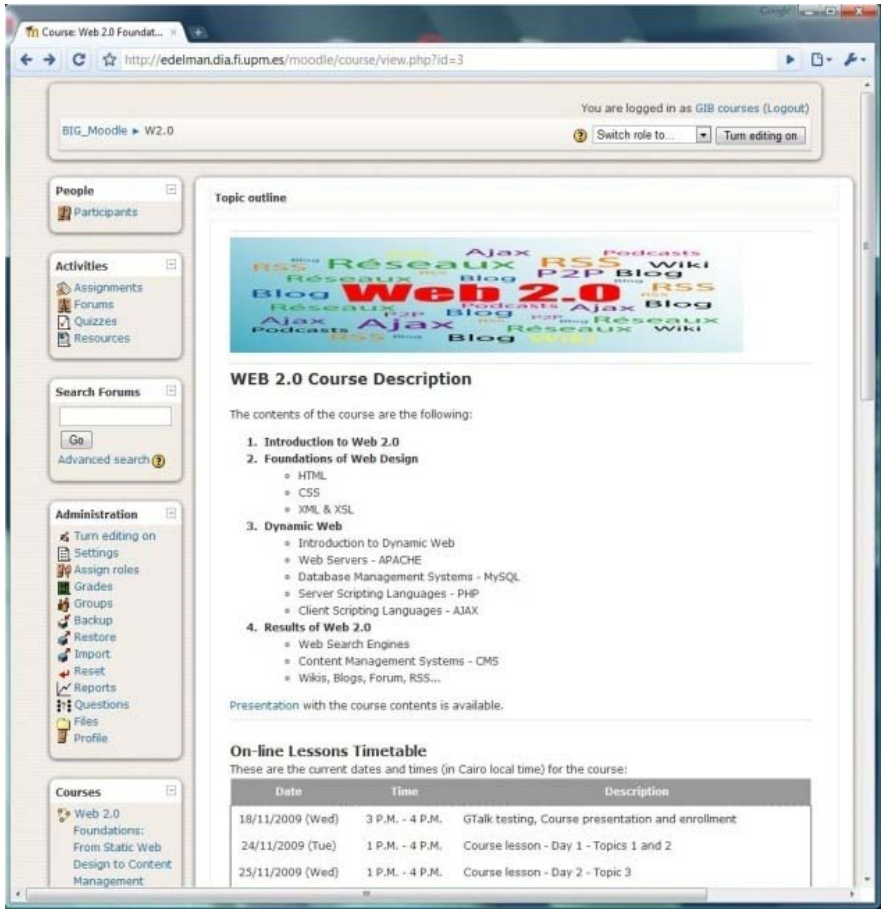
Large collection of BMI systems developed at the UPM group

- ONTOFUSION, for database integration.
- Brokerage Service, applied for Mobility and Training.
- OntoDataClean, for data mining.
- Inventory of resources, for storing and accessing remote software tools.
- Protocol manager, multimedia tools for visualizing clinical practice guidelines and protocols.
- Vocabulary server, for managing biomedical ontologies and terminologies.
- Mapping tool, for semantic integration of terminologies and ontologies.
- SIAC, an expert system for medical emergency management.
- UPM-Text miner, tools for text mining.
- Clinical trials manager, for managing clinical trials on cancer.
- Gene-Pdf, to convert contents of pdf files containing genetic information.
- Open PACS builder, a system for building small open-source PACS.
- Numerous Web services for image processing, data management and text and data mining.
- A software tool for remote collaborative work.
- Peer to peer image exchange tool.
- Various visualization tools for medical images.
- Geographical Information System, for planning pathways.
- BIRI, an automated inventory of bioinformatics resources
- DNAPubFinder
- PrimerXtractor, for bioinformatics

On-site teaching activities

One of the GIB's members teaching Web technologies to young students in Burundi (left)and showing how to carry out the experiments described in this talk to a local physician (right)

e-learning courses

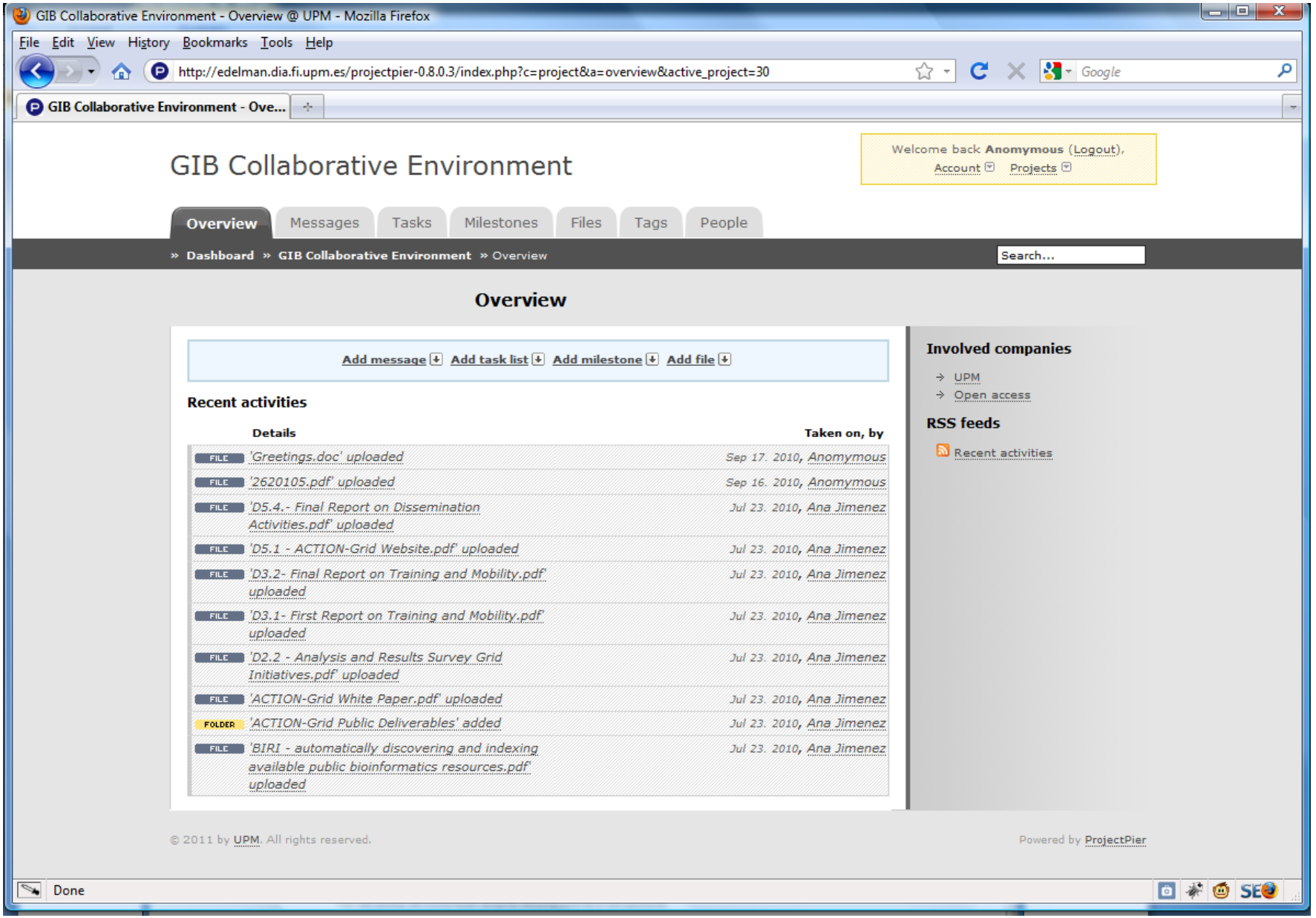


- Introduction: medical data, information and knowledge management
- Data mining: methods and tools
- Biomedical data integration: techniques for structured and non-structured sources
- Text mining
- Biomedical ontologies, terminologies and standards
- HL7

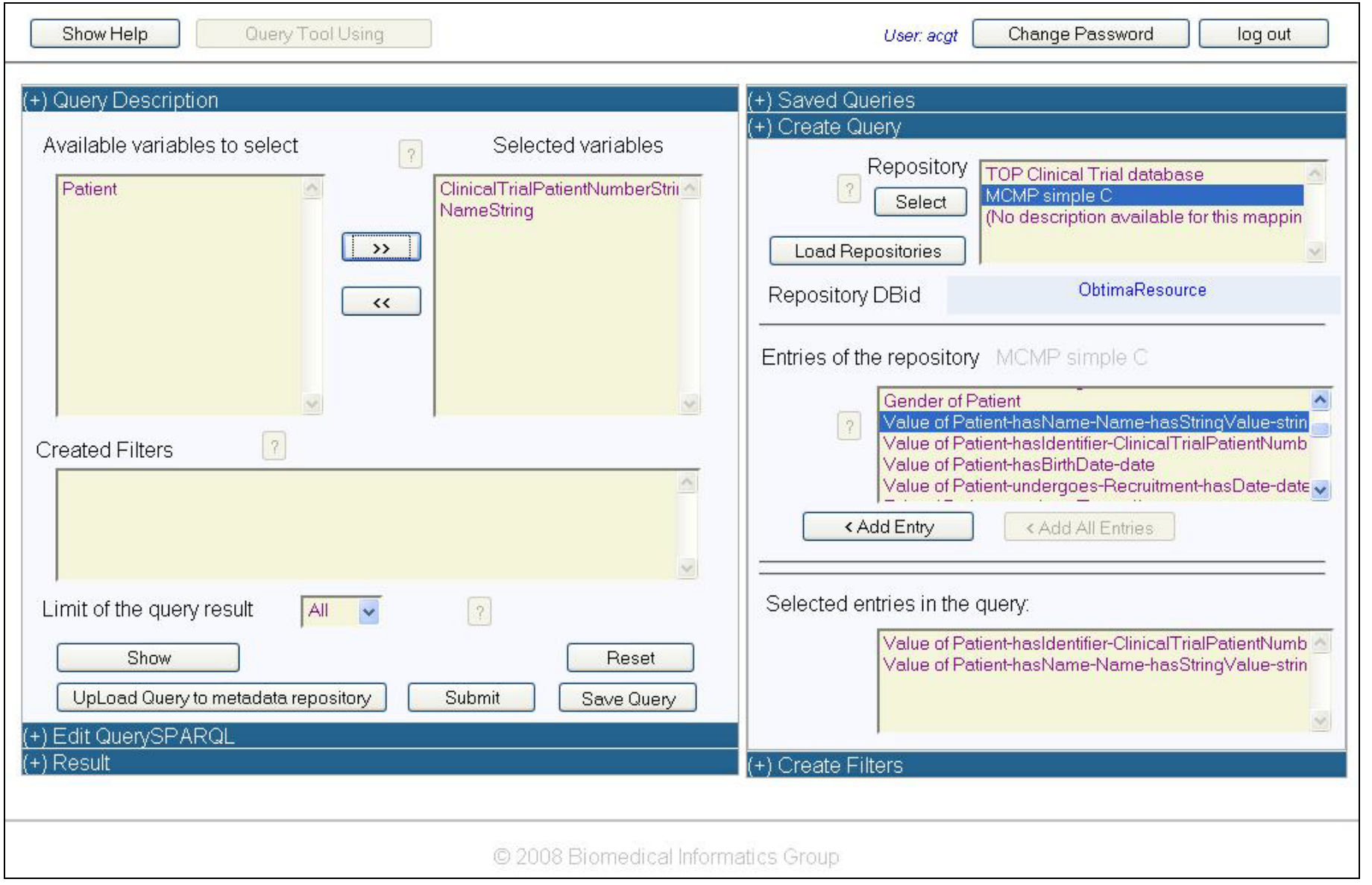
Catalogue of courses

Name of the course	*Leaders*
Short introduction course on Biomedical informatics	UPM
Hands-on training on OpenMRS and open source electronic medical records	ITI
Building clinical and research databases	UPM
Courses on pure computing topics	UPM
The WHO Reproductive Health Library: philosophy and use in practice	WHO-RHR
Evidence based reproductive health: How to read and write scientific papers	WHO-RHR
Knowledge Transfer and Exchange	WHO-RHR
Linking Public Health data with Geographical Information Systems	UPM
Research in Maternal and Perinatal Health	WHO-RHR
Managing research and data for HIV/AIDS with Medical Information Systems and Web technologies (to be created)	UniGE, UPM
Quantitative vs. Qualitative Research and Data collection Instruments	WHO-RHR
Research in reproductive health: design and methodologies	WHO-RHR, UniGe
An introduction course to environmental health	UniGE, ITM
Hospital Information Systems	UPM
Image processing and analysis (including PACS and virtual surgical planning)	UPM
Biomedical signal processing	UPM
Medical anthropology	All partners
Biomedical Informatics standards (vocabularies, terminologies and ontologies)	UPM
Clinical decision making and evidence-based medicine	ITM
Online learning methods for basic statistics	ITM
Literature search and critical reading	ITM
Introductory course on Health economics	UniGE, WHO-RHR, ITM
Data Mining in biomedicine	UPM
Text mining and information retrieval in biomedicine	UPM
Decision Support Systems	UPM
Methods for evaluation in the health sciences	UniGE, UPM
Web 2.0 and Semantic Web	UPM
Case studies	All partners
Application of informatics tools for teaching basic (e.g., anatomy, physiology) and clinical medicine (e.g., pathology	All partners

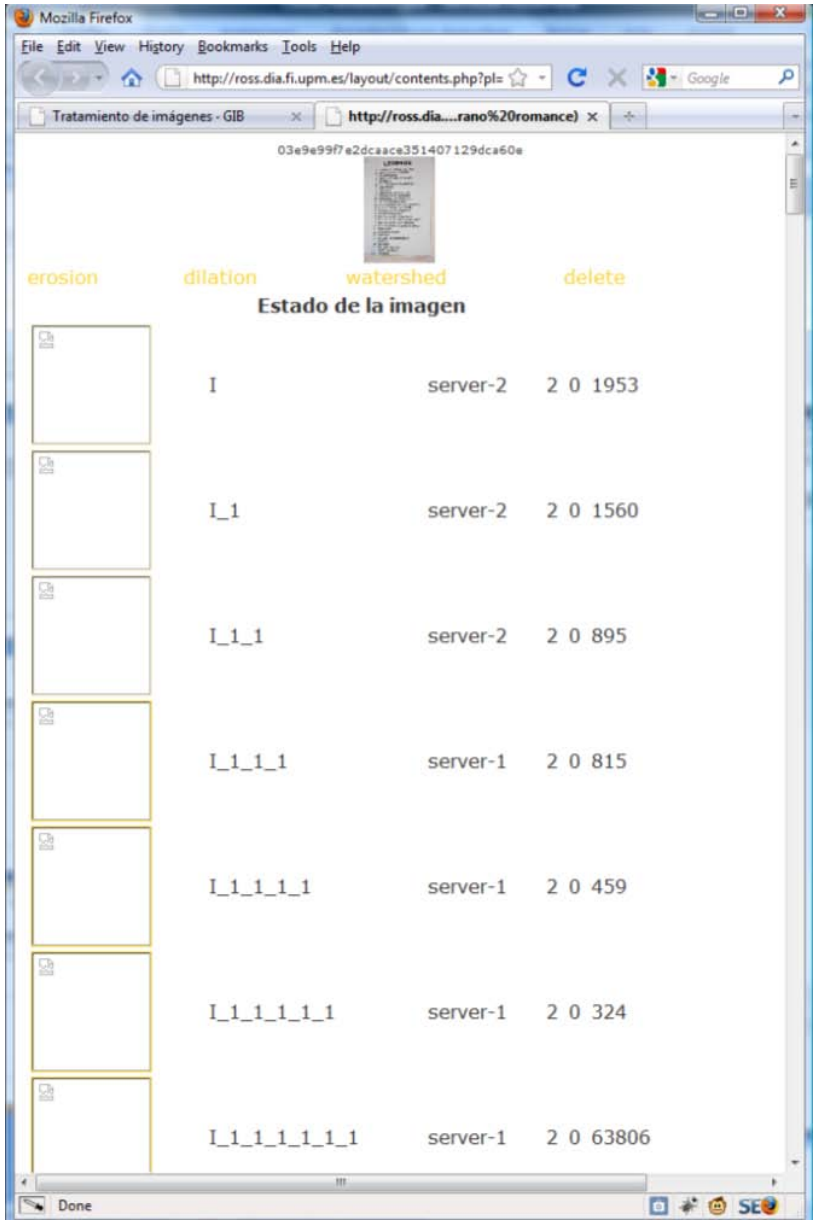
Collaborative environment

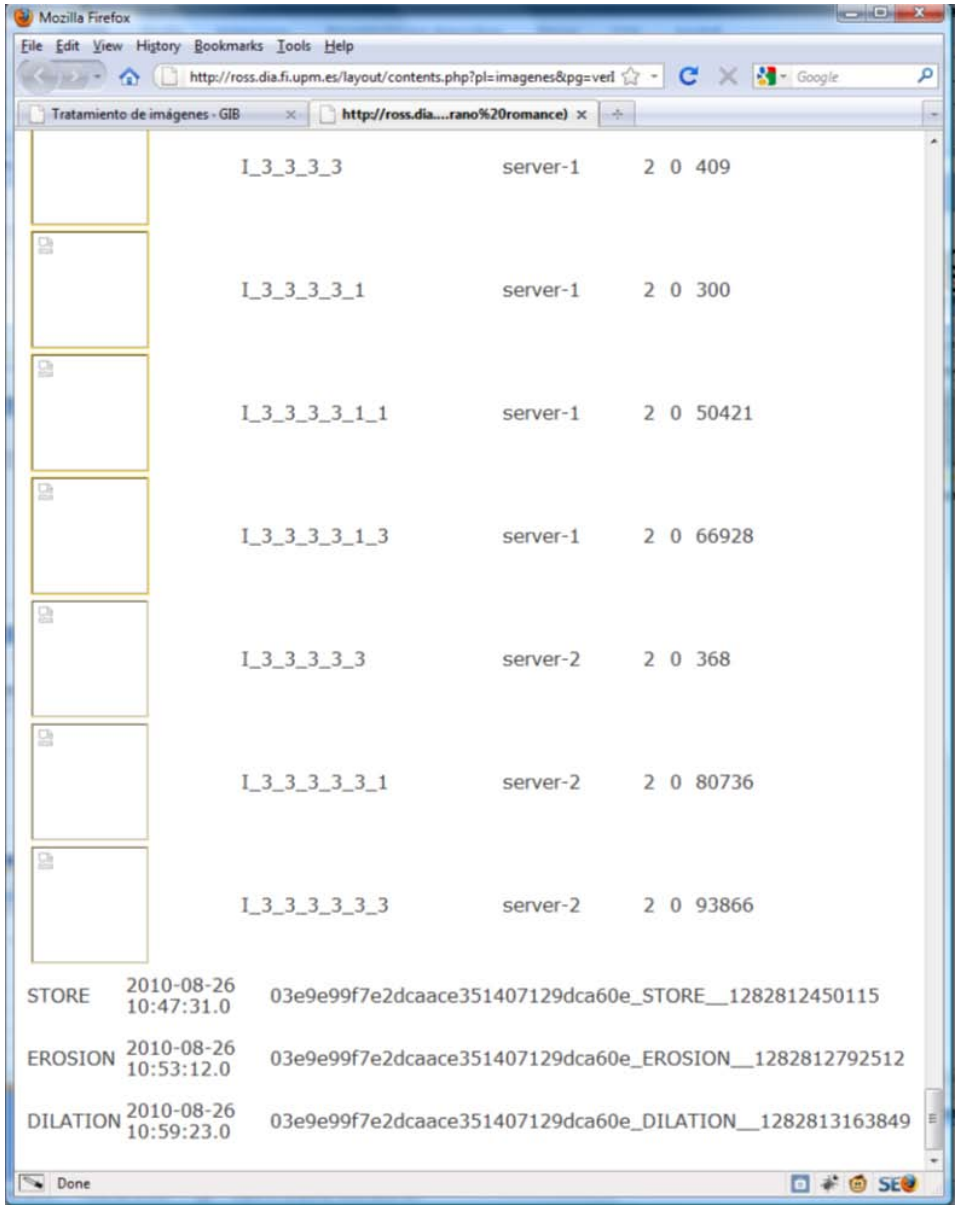


Semantic Integration (databases from Egypt accessed in Spain from Burundi)

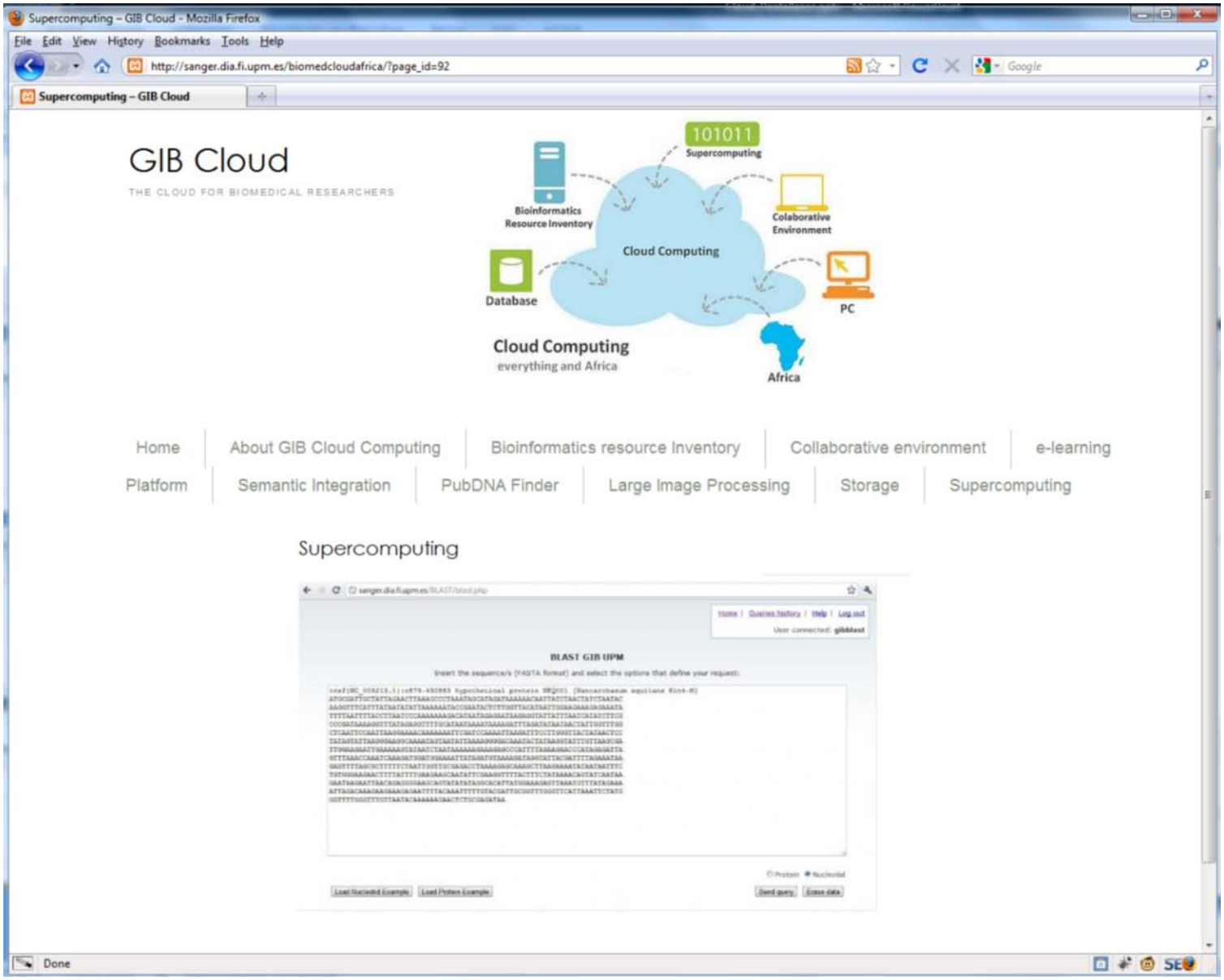


Large Image Processing from Burundi and Egypt





Access to the UPM supercomputer



Cork, Sept 2011, Victor Maojo

Final remarks

Three main ideas:

1. A global open (concerning both data and services) approach to biomedicine, from nanomedicine to public health
2. New approaches for building visual, graphical biomedical ontologies, far from current approaches
3. Extending the use of these (and previous) techniques to developing countries