Using Web Services for Linking Genomic Data to Medical Information Systems

V. Maojo, J. Crespo, G. de la Calle, J. Barreiro, M. Garcia-Remesal
Biomedical Informatics Group, Artificial Intelligence Lab, Universidad Politécnica de Madrid, Spain

Summary

Objectives: To develop a new perspective for biomedical information systems, regarding the introduction of ideas, methods and tools related to the new scenario of genomic medicine.

Methods: Technological aspects related to the analysis and integration of heterogeneous clinical and genomic data include mapping clinical and genetic concepts, potential future standards or the development of integrated biomedical ontologies. In this *clinicomics* scenario, we describe the use of Web services technologies to improve access to and integrate different information sources. We give a concrete example of the use of Web services technologies: the OntoFusion project.

Results: Web services provide new biomedical informatics (BMI) approaches related to genomic medicine. Customized workflows will aid research tasks by linking heterogeneous Web services. Two significant examples of these European Commission-funded efforts are the INFOBIOMED Network of Excellence and the Advancing Clinico-Genomic Trials on Cancer (ACGT) integrated project.

Conclusions: Supplying medical researchers and practitioners with omics data and biologists with clinical datasets can help to develop genomic medicine. BMI is contributing by providing the informatics methods and technological infrastructure needed for these collaborative efforts.

Keywords

Biomedical informatics, Web services, genomic medicine, hospital information systems

Methods Inf Med 2007; 46: 484–492

doi:10.1160/ME9056

Introduction

The successes of the Human Genome Project and other omics-related projects have resulted in a number of promises and challenges, including new approaches to basic medical science and healthcare, labeled generically *genomic medicine*. We and others have already addressed this issue elsewhere [1-3]. The introduction of other information, such as microarray data, biomarkers, single nucleotide polymorphisms (SNPs), etc., promises to provide physicians with comprehensive genetic information, advancing the scientific knowledge of physiological and pathological processes. Diseases such as cancer, diabetes, rare diseases and many others are now being explained from different perspectives, one of which is genetic information. With all these new data, prevention will take a more active role in public health [4-6].

To advance towards genomic medicine, medical informatics (MI) and bioinformatics (BI) researchers are beginning to exchange ideas at the intersection of the disciplines, e.g. in areas like systems biology and physiology. At a European level, the European Commission (EC) has launched a number of initiatives over the last few years to develop new BMI projects and move towards genomic medicine. These ideas were officially launched in November 2001 at a conference held in Brussels, entitled "Synergy between Research in Medical Informatics, Bioinformatics and Neuroinformatics. Knowledge Empowering Individualized Healthcare and Well-Being" [6]. Over these five years the EC has strongly supported research projects, conferences and several initiatives to create a new discipline that aims to link MI and BI, named biomedical informatics (BMI).

In this scenario, many collaborative actions are being proposed at an international level, particularly in Europe. They include topics such as biobanks, linking metabolic pathway representations with clinical data or creating new data models and standards for clinico-genomics research, medical records or hospital information systems (HIS) [7]. The goal is not to "deconstruct" the steps of clinical practice to provide the computer tools that health professionals need, but to shift clinical practice towards new approaches, founded on the underlying basic – biological – science. To test the hypothesis that can be generated in this scenario, researchers will need new models of genotype-phenotype links and more sophisticated approaches to classical clinical trials, including omics information.

In this paper we present an analysis of different issues related to linking clinical and genomic data in the framework of classical information systems – e.g., HIS. We analyze the use of Web services for this purpose, including some specific examples that we have implemented.

Accessing Genomic Information for Biomedical Research and Practice

The Human Genome Project favored a new trend in scientific research by establishing collaborative links among remote institutions that shared software tools and data across remote sites. By contrast, most clinical information databases are not openly accessible due to confidentiality and privacy issues regarding restricted patient information. For personalized medicine, both ap-

proaches must be combined. Whereas information should be as open as possible, confidential data must be protected.

For research in the area of genomic medicine, biomedical researchers and practitioners need to access and analyze data that may be only partially available. When clinical data is needed, confidentiality issues arise. This data can be necessary for many aspects related not only to patient care and management but also to research. For instance, over the last few years BI professionals have realized that to extend their molecular biology or systems biology results to physiology and pathology they will need to analyze and test their results using clinical datasets. Linking clinical and genomic information raises a number of research issues, such as:

- Collecting and retrieving information from remote heterogeneous sources
- Interoperability of clinical and genomic systems
- New standards for representing and exchanging data
- Mapping clinical and genomic data
- Development of new integrative ontologies, containing clinical and genomic concepts (or classes)
- Models for representing clinical and genomic data and knowledge in a genetic medical record
- Integrative genomic approaches at the testbed: diagnosis, prognosis and management (e.g., the Boston-based National Center for Biomedical Computing's Informatics for Integrating Biology and the Bedside (I2B2))
- Security and data confidentiality
- Data mining of clinico-genomic datasets: dealing with semantic heterogeneity, uncertainty and noise
- Building information and decision support systems to help professionals that are not familiar with new types of information (e.g., physicians that might not be familiar with SNP representation or molecular biologists that have not handled patient datasets before)
- Training in a new environment involving many disciplines: mathematics, computer science, medicine, biology, public health, and so on

These are mostly engineering challenges, but there are also some complex scientific questions to be addressed. The integration of omics information into different clinical systems, such as HIS or the electronic medical record, raises a number of issues that should be at the top of the future BMI research and development agenda. The integration of omics data into HIS, for instance, involves many of the challenges listed above.

At the time of writing, over 850 biological (e.g., genomic, SNPs, proteomic and other) databases were publicly open to researchers and other users. These databases are the result of many biological research projects that have delivered a large amount of heterogeneous information about genes, proteins and genetic diseases – e.g., nucleotide polymorphisms, gene mutations and others. Public databases are created and maintained by different institutions and research centers. Public databases often include related data types – e.g., Prosite, Swiss-Prot, and Protein Data Bank (PDB) include protein data. In other cases, organizations store their own information – e.g., SNP and mutation databases –, but they usually focus on just one topic. There are a number of different initiatives aiming to integrate information from Web-based public databases and private, local databases to facilitate biomedical research. Whether or not the publicly available information is integrated will have a significant impact on future clinical applications of genomic research.

Let us look at an example. Some biomedical researchers or clinicians need to gather specific genetic information that is stored somewhere on the Internet. These researchers should be aware that the information they are looking for may be available in many remote databases. Therefore, they need to identify where this data is located. Once researchers have located the names and Uniform Resource Locators (URLs) of databases that are of interest, they then access the URL and find a Web page, like the one shown in Figure 1 (hosted at the European Bioinformatics Institute (EBI). This institution maintains several databases and provides direct links to others.

After the researchers have accessed this Web page, they will select the database of their preference and will specify a definite query. The system will display the results of the query and the researchers will use this information.

This example is composed of several problems. Researchers must locate the information that they need from over 850 databases. They must know the database contents and how to access them. They must know how to navigate the Web page, make a specific query (e.g., using some controlled vocabulary, specific commands, coding, etc.) and retrieve the information needed. Results must be formatted adequately and their contents must be understood by the users. Finally, this information should be combined with previous knowledge to establish hypotheses and clinical diagnoses, advance biomedical research, and so on. This kind of approach is slow, tedious and unaffordable if the researcher needs, for instance, to access a large number of databases. Therefore, institutions are looking forward to using new technologies to enhance data access and retrieval.

Accessing Remote Heterogeneous Information: Web and Grid Services Technologies

Different Web technologies have been proposed for accessing and sharing remote heterogeneous information. One of the most significant is the development of Web services. Web services are a new programming trend for distributed systems that conform to a stateless programming model. Web services are based on service-oriented architectures using established Web standards such as Hypertext Transfer Protocol (HTTP) and eXtensible Markup Language (XML). Different protocols have been introduced, such as: 1) Web Services Description Language (WSDL) [8] for describing the services, 2) Simple Object Access Protocol (SOAP) [9] for defining the format of the data messages exchanged, or 3) Universal

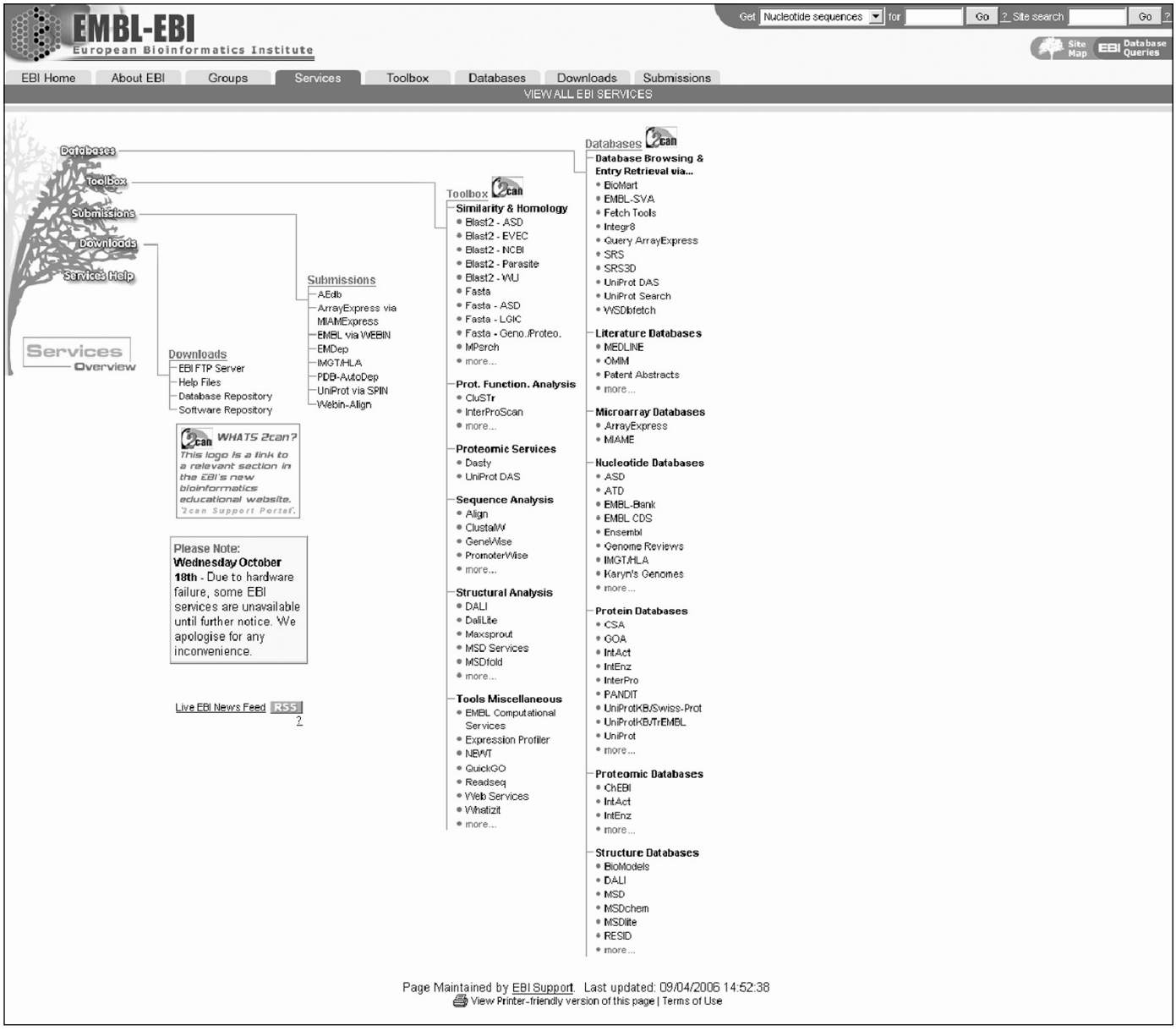

Fig. 1 An example of the EBI (European Bioinformatics Institute) public databases access interface

Description, Discovery, and Integration (UDDI) [10] for registering and discovering services. These protocols are all based on XML. Web services are completely platform-independent since they are based on XML. This feature makes the interoperability among services very high. Services can be deployed and invoked from any platform or architecture. Although they provide an underlying basis for designing Web architectures, new standards for dealing with semantic heterogeneity are still under debate.

Web services have become popular in areas of the life sciences over the last few years. Five years ago, Web services were an incipient technology aimed at improving distributed programming models for sharing resources among organizations. An article published in Nature [11] suggested that new Web technologies would soon be available to create the idea of a "bioinformatics nation", i.e. a common place where researchers could easily share different BI systems, tools and data, without having to access and be familiar with the different features of each site. For this vision to materialize, new integrating technologies and standards would be required. Without such technologies, biomedical informaticians could create thousands of databases and information sites, whose integrated access would be almost impossible. This need is even greater within the framework of personalized and genomic medicine, where clinical and omics – e.g., genomics, proteomics, metabolomics, and many other – data and systems should be integrated.

Following this new vision, researchers have used Web services for many different applications. Many systems are currently available in the BMI area. In the following, we mention a few significant and recent examples of the successful use of Web services for deploying distributed systems in the biomedical area.

In the genomic area, researchers at INRA (the National Institute for Agronomic Research in France) have developed AGMIAL [12], an open-source platform based on Web services for annotating bacterial genomes. Its architecture is composed of two main modules: the Protein Analysis Module (PAM) and the Contig Analysis Module (CAM). They are used to annotate protein and genomic sequences, respectively, and both provide support for storing and communication.

Other projects like NemaFootPrinter [13] or the Center for Biological Sequence Analysis' (CBS) Genome Atlas Database [14] apply Web services to access genome annotation databases. At the EBI [15] many resources are accessible by means of Web services. The services offer diverse possibilities for the use of analysis tools, homology searches, multiple alignment, structural analysis, text mining or data retrieval. All services are implemented following common Web services standards. Clients written in Perl and Java are freely available for most. Many other institutions have migrated their systems to Web services to facilitate data search and access. Anyway, the interoperability and "orchestration" of heterogeneous systems at different institutions is still under debate.

The reason for having to integrate different Web services from disparate organizations is clear. Isolated services are often insufficient for solving complex problems and the growing needs of BMI researchers. The possibility of combining these services in realistic workflows, adapted to the exact demands of researchers, from the combination of previously available is a new challenge. The classic problem of "reinventing the wheel" in BMI, i.e. developing systems that have been already developed elsewhere, appears once more. New approaches have been considered to deal with the need of reusing and orchestrating "services".

A good example is Taverna [16], developed for the myGrid Project [17] and funded by the European Commission. Taverna was designed to facilitate the integration of distributed resources (usually Web services) into workflows. Workflows can be built and manually executed on different system environments. Once a workflow is defined, any user can use it. As reported by its developers, around 3000 services were registered for use in the system in March 2006 [16].

Numerous languages and standards have appeared to support Web services-based workflow definition and execution. XLANG (Web services for Business Process Design) [18] was a pioneering initiative proposed by *Microsoft* as an extension of WSDL to provide the model for services orchestration and collaboration among orchestrations. BPML (Business Process Modeling Language) [19] is another open specification that defines a formal model for designing executable and business processes. Both XLANG and BPML are based on an explicit calculus theory. WSFL (Web Services Flow Language) [20] was launched by *IBM* to combine existing services and develop new ones. It defines two types of compositions: *flow models* to specify the usage of services and *global models* to establish interactions among services. BPEL4WS (Business Process Execution Language for Web services) [21] is another, *BEA*, *IBM*, *Microsoft*, *SAP AG* and *Siebel Systems* initiative that covers the specifications proposed by XLANG and WSFL. All these languages are based on common Web services concepts and standards, like WSDL, SOAP and UDDI. On the other hand, there are also radically different approaches to these "de facto" standards, such as ebXML-BPSS (electronic Business using XML – Business Process Specification Schema) [22] developed by *OASIS* or SCUFL (Simple Conceptual Unified Flow Language) [23], the workflow language defined for the Taverna project.

BioMOBY [24] is another example of an open source BI system using Web services. BioMOBY focuses on the development of innovative techniques for improving services registration and discovery for sharing biological data. Based on a decentralized architecture, BioMOBY implements its own specific central registry. This registry does not conform to the UDDI specification because of the standard's limitations. The central registry stores information about each service, such as the service name, service type, inputs, outputs or services provider's Uniform Resource Identifier (URI). A number of projects are currently using BioMOBY. In Spain, for instance, the INB (National Institute for Bioinformatics), a network of selected BI research institutions, has adopted BioMOBY to store and link different systems and tools [25].

Another interesting application is Gbrowse Moby [26]. This is a Web portal that gives access to databases and analytical tools. Users can explore and navigate the contents of the registry using a Web interface especially designed for non-informaticians. This portal is based on a novel semantic Web service browsing approach for using bioinformatics data and services, designed to create simple and linear workflows. Although it has a number of weaknesses as regards the construction of complex pipelines, it is a good example of the benefits that could be gained by applying semantic Web services technologies.

Recently, efforts are being made to fuse components of BioMOBY and Taverna [27]. For instance, a new plug-in for the Taverna workbench gives access to the BioMOBY central registry. All service descriptions retrieved from the BioMOBY registry can be used to define new Taverna workflows. Such descriptions are specified by means of an ontology, called the BioMOBY Object Ontology.

Web services have other important performance-, availability- and reliability-related snags that should be considered before they are adopted by research institutions. For instance, since they are distributed over the Internet and offered by different companies and institutions, a service may not be available at the precise moment that it is needed. Another problem is services identification and description. Even though there are services registration and discovery protocols, definitions are purely syntactic. Services providers often do not specify any kind of metadata to describe services functionality. Manually combining services without this extra (semantic) information

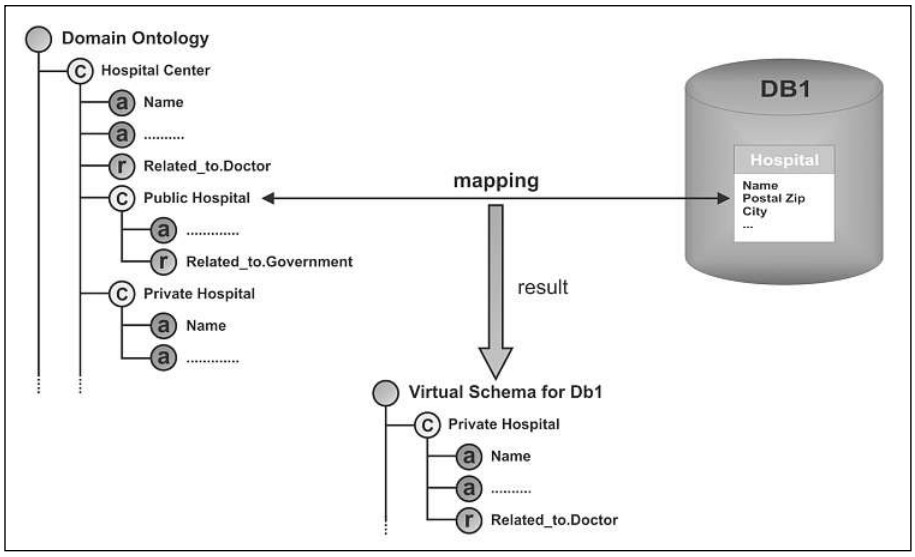

Fig. 2 Elements of the physical schema of the database are mapped to elements of the domain ontology.

can be a highly time-consuming task for researchers. Research on developing new workflows technologies can automate Web services use by specifying the expected inputs and the outputs and the concrete workflow needed for each application. A properly annotated resourceome [28], containing semantic information about all the available services, needs to be built to achieve this objective. Web services and BMI tool ontologies are needed to move in this direction.

We briefly describe a specific system that we developed to access and integrate remote services from heterogeneous sources, including both private and public clinical and biological databases. Its latest version has migrated to Web services technologies.

OntoFusion

OntoFusion [29] is a database integration system that provides homogeneous access to clinical and omics databases available over the Internet. Three kinds of sources have been integrated in the system: i) private databases, ii) public databases like OMIM (Online Mendelian Inheritance in Man), PubMed, PDB or Prosite, and iii) vocabularies and terminologies databases like UMLS (Unified Medical Language System), GO (Gene Ontology) or HGNC (Human Gene Nomenclature Committee). Information contained in these sources is often disparate and heterogeneous. We introduce domain ontologies to address this semantic heterogeneity. Before entering a new source to the system, we enact mapping and unification processes to associate real information with concepts and terms defined in the reference ontology.

OntoFusion follows a hybrid approach – based on ontologies for mapping databases and a query translation –, using virtual schemas to represent different sources.

Figure 2 gives an example of the semantic approach used in OntoFusion. A database (in this case, patient information from a clinical information system) represented as an ontology is mapped to a domain ontology. If the elements of the ontology representing the database cannot be mapped to any domain ontology, then the database schema can be modified. This kind of approach ensures semantic consistency and facilitates integration, since once the mapping is complete the system can automatically integrate the different "virtual" schemas, generated from the classes, relations and attributes identified in the domain ontology.

Taking this kind of approach the system can be used to integrate both private and public databases. The example below illustrates the results of a query within the OntoFusion system, used for research carried out in the context of the ACGT project [30], where we use this technology to create new clinico-genomic trial models [31]. The results retrieved from different public databases are integrated and displayed as elements of an ontology that users can browse.

OntoFusion Plus is an evolved version of the previous system reported elsewhere [29]. The improvements added to this new version focus on two topics: Web services and OWL (Web Ontology Language) [32]. Originally, OntoFusion was implemented using a multiagent architecture, based on the JADE (Java Agent DEvelopment Framework) platform [33]. A mediator module was responsible for distributing the queries among the virtual repositories and collecting the results. The query language used was RDQL (RDF Data Query Language) [34] and the virtual schemas and results were defined in DAML+OIL (DARPA Agent Markup Language + Ontology Interchange Language) [35]. Several public databases and vocabularies, such as OMIM, Prosite, PDB, GO or UMLS, were successfully integrated within OntoFusion.

After testing the system, we realized that most of the technological problems detected in our tests were caused by the agent platform. Therefore, a new architecture was needed. At that time, Web services were emerging as a standard technology for communicating and accessing distributed resources, and we decided to migrate to that infrastructure. Similarly, OWL has become the "de facto" standard language for knowledge representation using ontologies. In the new version, all the information previously represented in DAML+OIL has been updated to OWL.

Architecture

The functionality of the new system is similar to the previous one, but the inbuilt features of Web services have provided different features and functionalities. An abstraction of this new architecture is represented below in Figure 4.

As mentioned above, the previous version of OntoFusion used a static software agent to represent each component of the original system (virtual schemas, databases, servers, and others). In the new model, each agent has been replaced by a Web service. OntoFusion Plus defines eight Web services:

- IWSManager: This service covers the other services. It is responsible for managing and coordinating the different clients dealing with Web services.
- AuthenticationServer: This service authenticates and validates users and sessions. It also registers and activates new system users.
- DriversServer: This service obtains and manages the drivers needed to physically access the integrated databases.
- SchemasServer: This service provides the information related to the *Virtual Schemas* registered within the system. These *Virtual Schemas* represent the conceptual structure of the sources.
- DBServer: This service provides the methods and functions needed to manage the physical databases.
- SearchsServer: This service provides information about the available resources registered in the system.
- CacheServer: This service aims to improve the performance of the system. The last queries are stored in a cache. If they are requested again, the results are returned without executing the query.
- QueryServer: This service executes the different queries across the integrated resources.

The move to this Web service-based scenario was not straightforward because Web services are *stateless*. Agents needed to remember some information about their configuration and state. To furnish the services with *memory*, it was decided to use a database where Web services consult or store all the required information. Other functionalities and improvements have been incorporated to the new version, such as user registration or security issues, which were not implemented in the original system.

In this architecture, a special Web services manager was defined to coordinate the other services. Any user can access the sys-

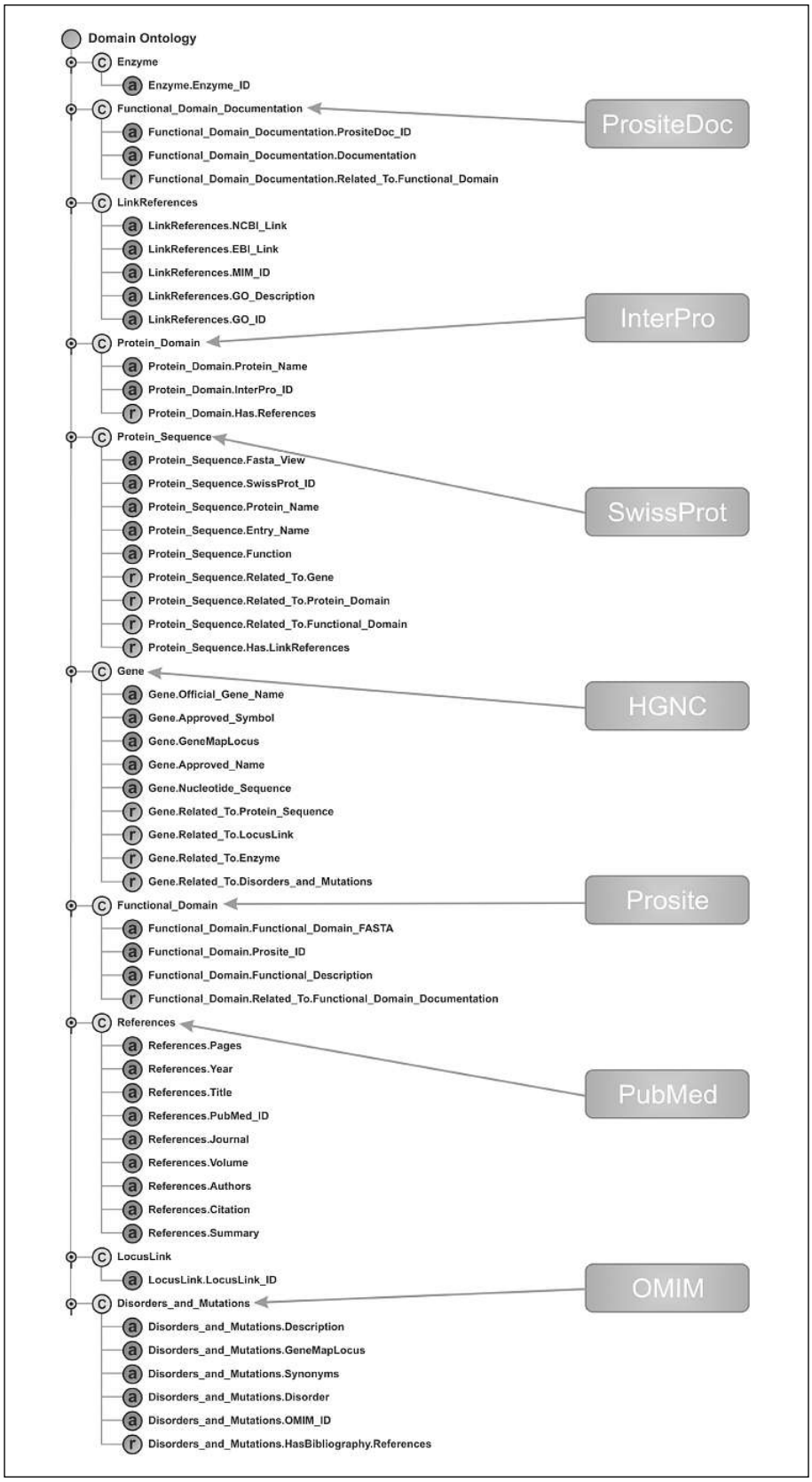

Fig. 3 An example of a query launched to OntoFusion. The system links information from different public databases and displays the results.

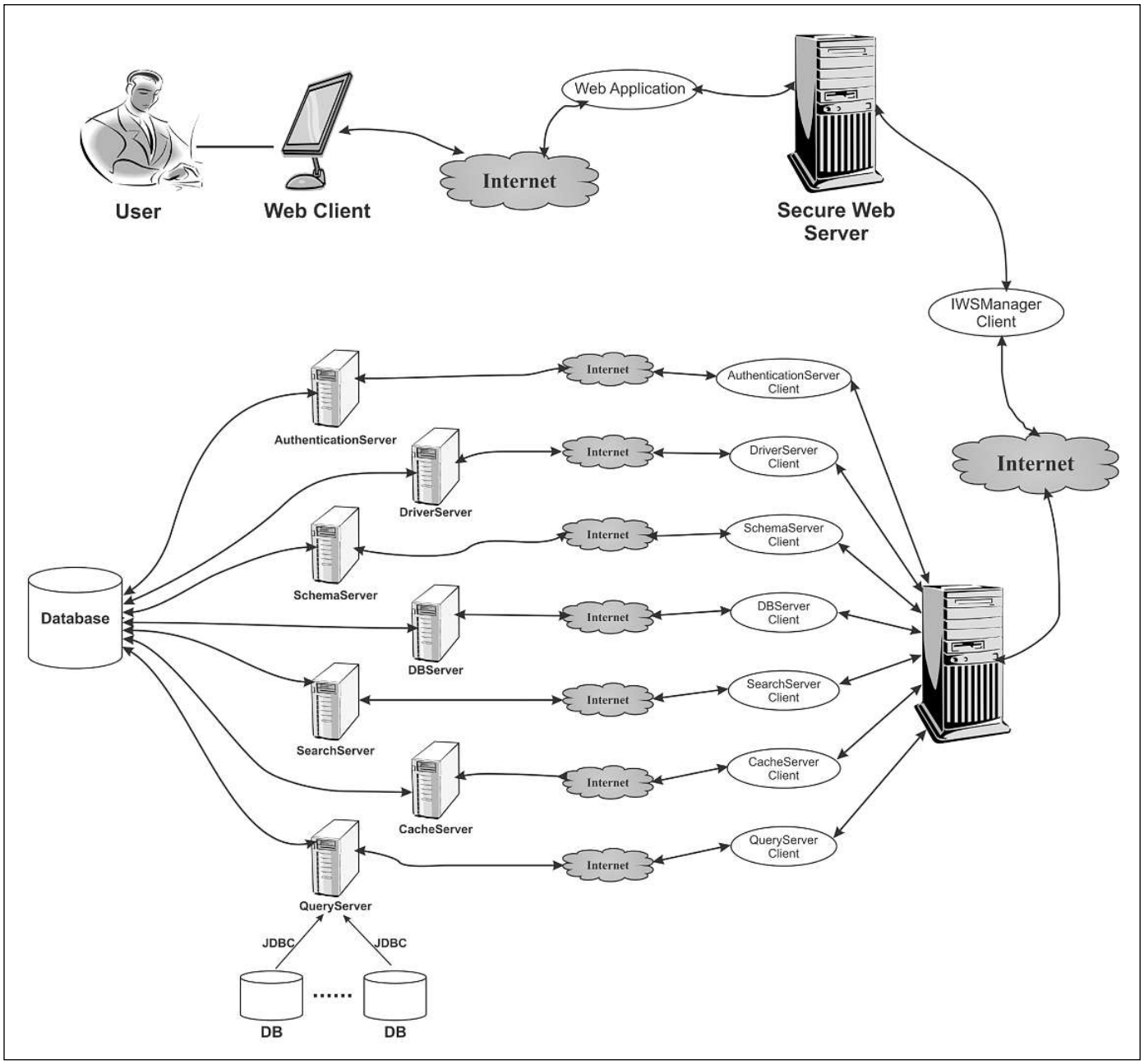

Fig. 4 OntoFusion Plus Web services-oriented architecture. It is based on a central Web service named IWSManager which coordinates the other seven defined services. Their functionalities, explained in the paper, range from security issues to database management.

tem from any location using an ordinary web browser. The requests are received by a web application running on a secure web server. This web application is responsible for dealing with the Web Services Manager (WSM) transparently from the user perspective. Depending on the information requested, the WSM distributes the requests across the appropriate Web services, returning the results to the web application in a standard manner. Usually the resources (e.g., services, databases, tools) are physically distributed across the Internet. Users need to know neither what services or databases are used nor how they are invoked.

Finally, we mention two different ongoing projects, in which we have participated. Both aim to link clinical and genomic information in the framework of personalized medicine. The final goal is to develop new BMI methods and tools to support biomedical research and practice.

The first is the EC-funded INFOBIOMED Network of Excellence [36]. It started up in 2004 with the objective of promoting BMI in Europe to support personalized healthcare. It has two main goals: the devel-

opment of new BMI technologies – e.g., database integration, ontologies, image processing, data mining – and their application to four clinical pilots. A number of pilots and tools have been developed around different topics related to heterogeneous, clinico-genomic integration within the INFOBIOMED framework. As an example, one of the clinical pilot projects targets hereditary colon cancer at an individual and family level. The responsible organization is the Danish HNPCC-register located at the Hvidovre Hospital in Copenhagen (Denmark). This public register has been built since 1991 and contains Danish families' genomic and epidemiological information. Early detection of Hereditary Non-Polyposis Colon Cancer (HNPCC) aims to increase patients' life expectancy. Traditionally, one of the major problems is that relatives usually reside at different places, which complicates analysis and diagnosis. This scenario suggested a new approach, exchanging genotype and phenotype data among institutions over the Internet. Several Danish molecular laboratories, genetic and surgery departments from different cities and institutions are involved in this pilot project. A Web-based application using HL7-based XML schemas, defined ad-hoc for exchanging data, has been deployed to access the integrated information. Data are coded using four types of XML messages that can describe pedigree information (family and individual), mutations and SNPs, and surveillance data.

Second, ACGT (Advancing Clinico-Genomics Trials on Cancer) [30] is another EC-funded project targeting the development of new methods and resources for cancer research. The objective of the project is to identify technological gaps in cancer research and create novel techniques for the early identification of the disease, as well as to improve research leading to new therapies. From a BMI perspective, one of ACGT's primary goals is to develop a biomedical Grid infrastructure that can be shared among European countries. With this infrastructure, partners are developing numerous Grid services – a concept closely related to Web services, needing more advanced standards and based on a Grid infrastructure. In this clinico-genomic trials scenario, heterogeneous data from remote, disparate sources will be integrated into a common virtual repository by means of a common interface. This repository combines information from clinical records together with medical images and both clinical and genetic patient data that can be used for clinical care and biomedical research, including in-silico simulation.

Conclusions

BMI faces many different information processing and management challenges at several levels: molecular, systems biology, cell, physiology, pathology and public health [3, 37]. These challenges are both scientific and technological. From a scientific perspective, there are still many gaps in our knowledge of many molecular, physiological and pathological issues, need to be filled to realize this vision. BMI can contribute to this research by providing tools for exchanging information, building in-silico simulations, mapping heterogeneous information, and so on. These tools have also helped to promote collaborative efforts among thousands of researchers, speeding up projects, such as the Human Genome Project, and contributing to an extraordinary improvement in patient care over the last few decades.

From a technological viewpoint, efforts in areas such as standards, Web services workflows, biomedical ontologies development, image processing, information retrieval and data mining have not only influenced biomedical research and practice but also pervaded other engineering disciplines. New professionals will need additional training in innovative aspects that have not been addressed before in classical training programs [38].

Hospital and medical information systems generally face a new situation. Genomic medicine sets new challenges as to gathering and managing different, remote and heterogeneous information (e.g., text, numerical, images, or sounds) that should be transparently integrated and managed by different categories of professionals [39, 40]. BMI should provide the methods and tools needed in this new scenario, providing scientific and technological support for biomedical research [41].

Acknowledgments

This work was funded by the INBIOMED research network (Ministry of Health, No. G03/160), the European Commission (FP6, IST Thematic Area) through the INFOBIOMED NoE (IST-2002–507585), the ACGT integrated project (FP6-2005-IST-026996) and the ONTOMINEBASE project (Ministry of Education and Science, Spain).

References

1. Collins FS. Shattuck Lecture – "Medical and Societal Consequences of the Human Genome Project". N Engl J Med 1999; 341: 28-37.
2. Kulikowski C. The Micro-Macro Spectrum of Medical Informatics. Challenges: From Molecular Medicine to Transforming Health Care in a Globalizing Society. Methods Inf Med 2001, 41: 20-24.
3. Martín F, Maojo V, López-Campos G. Integrating genomics into health information systems. Methods Inf Med 2002; 41 (1): 25-30.
4. McKusick VA. The anatomy of the human genome: a neo-Vesalian basis for medicine in the 21st century. JAMA 2001; 286 (18): 2289-2295.
5. Subramanian G, et al. Implications of the human genome for understanding human biology and medicine. JAMA 2001; 286 (18): 2296-2307.
6. Maojo V, Martin-Sanchez F. Bioinformatics: Towards New Directions for Public Health. Methods Inf Med 2004; 43: 208-214.
7. Sax U, Schmidt S. Integration of genomic data in Electronic Health Records – opportunities and dilemmas. Methods Inf Med 2005; 44 (4): 546-550.
8. Web Services Description Language (WSDL) 1.1 (online). Mar 15, 2001 (cited Feb 15, 2007). Available from: URL: http://www.w3.org/TR/wsdl
9. Simple Object Access Protocol (SOAP) 1.1 (online). June 24, 2003 (cited Feb 15, 2007). Available from: URL:http://www.w3.org/TR/soap/
10. Universal Description, Discovery and Integration (UDDI) (online). 2006 (cited Feb 15, 2007). Available from: URL: http://www.uddi.org/
11. Stein L. Creating a bioinformatics nation. Nature 2002; 417 (6885): 119-120.
12. Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessieres P, Gibrat JF. AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. Nucleic Acids Res 2006; 34 (12): 3533-3345.
13. Rambaldi D, Guffanti A, Morandi P, Cassata G. NemaFootPrinter: a web based software for the identification of conserved non-coding genome sequence regions between C. elegans and C. briggsae. BMC Bioinformatics 2005; 6 (Suppl 4): S22.
14. Hallin PF, Ussery DW. CBS Genome Atlas Database: a dynamic storage for bioinformatic results and sequence data. Bioinformatics 2004; 20 (18): 3682-3686. Epub July 15, 2004.

15. European Bioinformatics Institute (EBI) (online). 2007 (cited Feb 15, 2007). Available from: URL: http://www.ebi.ac.uk
16. Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T. Taverna: a tool for building and running workflows of services. Nucleic Acids Res. 2006; 34 (Web Server issue): W729-732.
17. myGrid Project (online). 2004 (cited Feb 22, 2007). Available from: URL: http://www.mygrid.org.uk/
18. Cover Pages (online). 2001 (cited Feb 22, 2007). Available from: URL: http://xml.coverpages.org/xlang.html
19. Business Process Management Initiative (BPMI.org) (online). Jan 02, 2007 (cited Feb 22, 2007). Available from: URL: http://www.bpmi.org/
20. Leymann F. Web Services Flow Language (WSFL 1.0) (online). May 2001 (cited Feb 22, 2007). Available from: URL: http://xml.coverpages.org/WSFL-Guide-200110.pdf
21. Business Process Execution Language for Web Services version 1.1 (BPEL) (online). Feb 08, 2007 (cited Feb 22, 2007). Available from: URL: http://www-128.ibm.com/developerworks/library/specification/ws-bpel/
22. Electronic Business using eXtensible Markup Language (ebXML) (online). 2006 (cited Feb 22, 2007). Available from: URL: http://www.ebxml.org/
23. Simple Conceptual Unified Flow Language (SCULF) (online). 2005 (cited Feb 22, 2007). Available from: URL: http://www.gridworkflow.org/snips/gridworkflow/space/SCUFL
24. Wilkinson MD, Links M. BioMOBY: an open source biological Web services proposal. Brief Bioinform 2002; 3 (4): 331-341.
25. Navas-Delgado I, del Rojano-Munoz M, Ramirez S, Perez AJ, Andres Leon E, Aldana-Montes JF, Trelles O. Intelligent client for integrating bioinformatics services. Bioinformatics 2006; 22 (1): 106-111. Epub Oct 27, 2005.
26. Wilkinson M. Gbrowse Moby: a Web-based browser for BioMoby Services. Source Code Biol Med 2006; 1: 4.
27. Kawas E, Senger M, Wilkinson MD. BioMoby extensions to the Taverna workflow management and enactment software. BMC Bioinformatics 2006; 7: 523.
28. Cannata N, Merelli E, Altman RB. Time to organize the bioinformatics resourceome. PLoS Comput Biol 2006; 2 (2): e20.
29. Alonso-Calvo R, Maojo V, Billhardt H, Martin-Sanchez F, Garcia-Remesal M, Perez-Rey D. An agent- and ontology-based system for integrating public gene, protein, and disease databases. J Biomed Inform 2007; 40 (1): 17-29.
30. Advancing Clinico Genomic Trials on Cancer (ACGT) (online). 2007 (cited Feb 22, 2007). Available from: URL: http://www.eu-acgt.org
31. Maojo V, Garcia-Remesal M, Billhardt H, Alonso-Calvo R, Perez-Rey D, Martin-Sanchez F. Designing new methodologies for integrating biomedical information in clinical trials. Methods Inf Med 2006; 45 (2): 180-185.
32. OWL Web Ontology Language Overview (online). Feb 10, 2004 (cited Feb 15, 2007). Available from: URL: http://www.w3.org/TR/owl-features/
33. Java Agent Development Framework (online). Feb 1, 2007 (cited Feb 15, 2007). Available from: URL: http://jade.tilab.com/
34. RDF Data Query Language (online). Jan 09, 2004 (cited Feb 15, 2007). Available from: URL: http://www.w3.org/Submission/RDQL/
35. The DARPA Agent Markup Language Homepage (online). Jan 13, 2006 (cited Feb 15, 2007). Available from: URL: http://www.daml.org/
36. The INFOBIOMED Network of Excellence Homepage (online). 2006 (cited Feb 22, 2007). Available from: URL: http://www.infobiomed.org
37. Pollard TD. The future of biomedical research: from the inventory of genes to understanding physiology and the molecular basis of disease. JAMA 2002; 287 (13): 1725-1727.
38. Rindfleisch TC, Brutlag DL. Directions for clinical research and genomic research into the next decade: implications for informatics. J Am Med Inform Assoc 1998; 5 (5):404-411.
39. Sikorski R, et al. Genomic Medicine, internet resources for medical genetics. JAMA 1997; 278: 1212-1213.
40. Brandt CA, Sun K, Charpentier P, Nadkarni PM. Integration of Web-based and PC-based clinical research databases. Methods Inf Med 2004; 43 (3): 287-295.
41. Kulikowski C. The Scope of Medical Informatics as Art and Science. Methods Inf Med 1996; 35: 185-188.

Correspondence to:
Prof. Victor Maojo
Facultad de Informatica
Universidad Politecnica de Madrid
Boadilla del Monte
28660 Madrid
Spain
E-mail: vmaojo@fi.upm.es