Current Computer-Aided Drug Design. 2013; 9(1): 108-117

Biomedical data integration in computational drug design and bioinformatics

José A. Seoane, Vanessa Aguiar-Pulido, Cristian R. Munteanu, Daniel Rivero, Juan R. Rabuñal, Julian Dorado and Alejandro Pazos

Department of Information and Communication Technologies, Computer Science School, University of A Coruña, Spain

Abstract
In recent years, in the post genomic era, more and more data is being generated by biological high throughput technologies, such as proteomics and transcriptomics. This omics data can be very useful, but the real challenge is to analyze all this data, as a whole, after integrating it. Biomedical data integration enables making queries to different, heterogeneous and distributed biomedical data sources. Data integration solutions can be very useful not only in the context of drug design, but also in biomedical information retrieval, clinical diagnosis, system biology, etc. In this review, we analyze the most common approaches to biomedical data integration, such as federated databases, data warehousing, multi-agent systems and semantic technology, as well as the solutions developed using these approaches in the past few years.

Keywords:
Data integration, data warehouse, federated database, multi-agent systems, ontologies, semantic web technologies, Biomedical data, computational drug design, bioinformatics.

INTRODUCTION

Exponential accumulation of biomedical data has the promise to improve the discovery of biomedical knowledge. This would permit the development of new generations of personalized biomedicine. The goal of a rational drug design is molecular modeling of certain diseases, prediction of specific molecules which interact with specific proteins, identification of a target genetic profile in population for a specific drug and absorption, distribution, metabolism and excretion prediction for a specific drug in silico. All these goals involve the rational use of biomedical knowledge, which is stored in distributed and heterogeneous biomedical data sources. Much of the currently available data is qualitative, noisy, inaccurate and incomplete. Effective data integration tools are a key component to reach this objective in system biology.

There is a lot of literature [1-19] about data integration in the area of biomedical informatics. In this review, technologies that can be used to solve data integration problems (data related to "omics") are described.

Integrating data consists, essentially, in making queries to different data sources, which can be either related databases or semi-structured data sources connected to a network. Reference [9] proposes an orthogonal organization for data integration. This organization refers to, on one hand, where the metadata is stored and, on the other hand, to how the data is represented and to the data models. There are some works that deal with the problem of omics integration with clinical data [20-23].

According to the 2010 update for the Bioinformatics Links Directory [24], there are nearly 1500 publicly web-accessible links including databases and web servers, that aim to collect, organize, visualize, integrate and analyze biological data. For a given task, researchers in the field of bioinformatics often need to consult numerous databases and web servers.

However, the integration of heterogeneous datasets from disparate databases associated with multiple web servers is daunting for researchers. It requires them to be proficient at computationally 'surfing' databases and web servers and algorithmically 'skimming' the requisite data. The challenge of decoding volumes of biological data from disparate sources underscores an imperative for greater data integration models.

OMICS

Determining the whole sequence of the human genome has been recognized as an important task more than two decades ago and was the precursor of genomics. From then on, human genetic diversity has been studied identifying different haplotypes present in the population, sequencing genomes of representative individuals.

In the post-genomic era, different methods have appeared allowing simultaneous evaluation of a great number of "transcriptomes" (messenger RNA) (transcriptomics), RNAi/miRNAs (interferomics and microRNAomics), proteins (proteomics), protein interactions (interactomics), DNA modifiers and chromatin (epigenomics) and metabolites (metabolomics), trying to contribute to describing and understanding the model of life.

The challenge in –omics research is to solve the problem of knowledge fragmentation, by integrating different heterogeneous information sources in a coherent entity. The integration of this data is essential for the omics community, as omic data is currently scattered in different databases and different formats all over the world. These formats have to be correctly integrated using, for this purpose, different techniques that will be described subsequently.

ONTOLOGIES

In the omics context, controlled vocabularies offer a form of data integration by enforcing naming conventions for data elements that lately appear in omics databases. In this type of technique, the integration of heterogeneous omics data sources is based mainly on ontologies [25-27].

The term ontology has its origin in classical philosophy. It is the name of the main branch of metaphysics and deals with being and existence. In the area of informatics, ontology can be defined as a data model that represents a set of concepts in a certain domain and the relationships among these concepts. An ontology is a specification of a conceptualization, being a conceptualization a simple and abstract model of the world. A specification is, thus, a declarative representation of a conceptualization in a specific format. It attempts to interpret knowledge in the same way a computer can process, without any ambiguity, and consequently encode concepts and relationships in a language that can be interpreted by a machine. The Ontology Web Language (OWL) is the standard recommended by the W3C to represent ontologies. OWL has a great interoperability with computers for the web, more than the Extensive Markup Language (XML) DARPA Agent Markup Language (DAML), Resource Description Language (RDF) or RDF Schema. OWL has three sublanguages: OWL Lite, OWL DL y OWL Full.

In the biomedical field, a large number of ontologies have been developed. The Unified Medical Language System (UMLS) is a great source of biomedical terminology, with a large vocabulary and international classifications. The National Center for Biomedical Ontology (NCBO) has built a library of biomedical ontologies known as Open Biomedical Ontologies (OBO) [28], which consists of more than 70 biomedical ontologies. The Ontology for Biomedical Investigations (OBI) has developed and integrated ontologies to describe life sciences and clinical research. The OBI Consortium is a member of the OBO Foundry. OBI is currently using the Basic Format Ontology (BFO), which is its high level ontology. The high level ontology captures most of the concepts that are basic to human understanding of the world. It also describes very general concepts which are identical in different domains. Ontologies have become crucial in the biomedical data integration field, especially since the semantic web irrupted into biomedical data annotation [16]. Promising applications of this can be found for drug design in reference [29].

Application of Ontologies in Data Integration

In the area of databases and information systems, ontologies allow integrating data from multiple heterogeneous data sources, transforming this data into a common representation and transmitting the knowledge to the software. Semantic heterogeneity has been identified as one of the most important challenges in data integration as it requires understanding relationships between the real world and the data, which are often very subjective. Ontologies propose a solution to the semantic heterogeneity problem, providing formal definitions of terms used in the data sources, as well as providing an implicit meaning to the relationships among the different data terminologies explicitly. Many organizations are now exploring the use of Semantic Web technologies in the hope of easing the cost of data integration [30]. There are several authors that point out that semantic technologies are the next step in data integration [25-27, 31-36].

LINK INTEGRATION APPROACHES

Most public databases are connected to internet and can be accessed from web pages. Most of these databases provide hypertext links to entries of other databases. In most cases, the accession number (stable identifiers specific to a database, which are generated when a new instance is added to it) is used to interconnect database records from web pages.

Normally, the mapping between two entries from different data sources has to be performed explicitly so that there exist links between them. For this reason, databases usually only provide links to the most used ones, as there are more than 500 databases of molecular biology and the organisms that maintain these databases have to take into account the most relevant databases to generate and maintain the links.

Integrating data through direct navigation removes the relational data model applying a model in which data sources are defined as sets of pages, which have interconnection points in order to establish relationships between them.

One of the major problems of this type of integration is that, frequently, information obtained in this phase may generate multiple links to other data sources to perform the following query. Web pages with links to other databases are the most common type of data integration, even though, usually, they are not as interconnected as desirable.

Link Integration Examples

This approach is probably one of the most popular and effective data integration in portals and keywords indexing systems. Examples of this approach are SRS, Entrez or Integr8. Goble [13] points out that in 2008, 40% of the EMBL-EBI traffic was from SRS queries. SRS [37, 38] (Sequence Retrieval System) was originally designed to access to biological sequence databases. Another important database for drug design developed by the EBI is ChEMBL [39], which contains binding, functional and ADMET information of a large collection of compounds. Currently one of the most important tools in biomedical research, providing access to more than 400 databases, NCBI Entrez [40] is one of the most widely used interfaces for information retrieval in biomedical databases. Entrez maps relationships between individual entries in public databases. In drug design is also very important the role of cheminformatic databases of the NCBI as PubChem [41], which is divided into tree databases, substance database, compound database and bioassay database. Integr8 [42] provides access to the full genomes and proteomes of more than 190 species, including general information, publications, etc. and performs statistical analyses. Finally, DisaseCard [43] is a web collaborative tool that aims to integrate genetic and biomedical information related with rare diseases.

DATA WAREHOUSING

Unlike the previous integration model, the data warehouse approach provides an integration model specifically designed for this purpose. Data warehousing consists in storing all the data of a certain type from different databases in only one large database, with a general schema, using several technologies [44, 45] (see Fig.1).

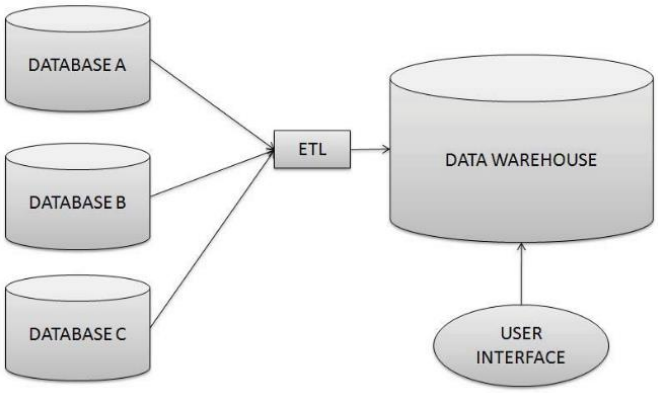

Fig. (1). Data warehouse approach.

The process of extracting the data from different data sources and storing it in the central large database is named ETL (Extract, Transform and Load). It can be divided into the extraction of the different data sources using a wrapper structure, the transformation to the data model of the warehouse and loading the data into the warehouse.

Data warehouse is considered to be a safe way to provide researchers a quick and effective way to answer to their queries. This is not trivial since the performance is normally one of the main requisites for biomedical researchers. However, this access system makes local control of data easy, which enables a better curation and filtering.

Creating a repository where all the data will be stored also presents a series of problems. The volume of the data is generally too large to be managed using a data warehouse. Updating data of a data warehouse may cause difficulties in its maintenance. Thus, problems may arise. For example, queries might be only as relevant as the last update of the data warehouse. Moreover, creating a global schema which reflects the reality of all the data types is complicated. Normally, the richness of each individual data source is lost when only the common elements are reflected in the global schema or, on the contrary, if all the data from all the data sources are maintained then the global model becomes unmanageable. Despite all these problems, data warehouse is usually the most adequate solution to create sets of highly curated data, focused on a specific research area.

Data Warehouse Examples

One of the main reasons that justify the Data Warehouse approach in the data integration problem is the quality of the data [46]. The curation and filtering features of this model make it the most suitable choice when data quality is essential.

Examples of data warehouse are the IGD (Integrated Genomic Database) [47], which stores human genetic, protein, structural and bibliographical data, the Adaptable Clinical Trial Database ACT/DB [48], which stores data related to cancer clinical trials using entity-attribute-value form. DataFoundry [49] uses a wrapper-mediator architecture to store in the warehouse data from PDB, SWISS-PROT, SCoP and dbEST. GIMS (Genome Information Management System) [50] is based on an object database that integrates genome sequence data with functional data on the transcriptome and on protein-protein interactions. The BIOMOLQUEST [49, 51] is a search engine that uses a data warehouse which integrates PDB, SWISS-PROT, ENZIME and CATH data.

Target Informatics Net (TINet) [52] is an approach for data integration that uses a federated model but with a judicious use of a local data warehouse. Ensembl [53] was considered by some authors as a data warehouse because it stores annotation data of the human genome with confirmed gene predictions that have been integrated with external data sources. The University of California Santa Cruz Genome Browser Database [53-56] is a warehouse that stores genome sequence data integrated with a large collection of related annotations.

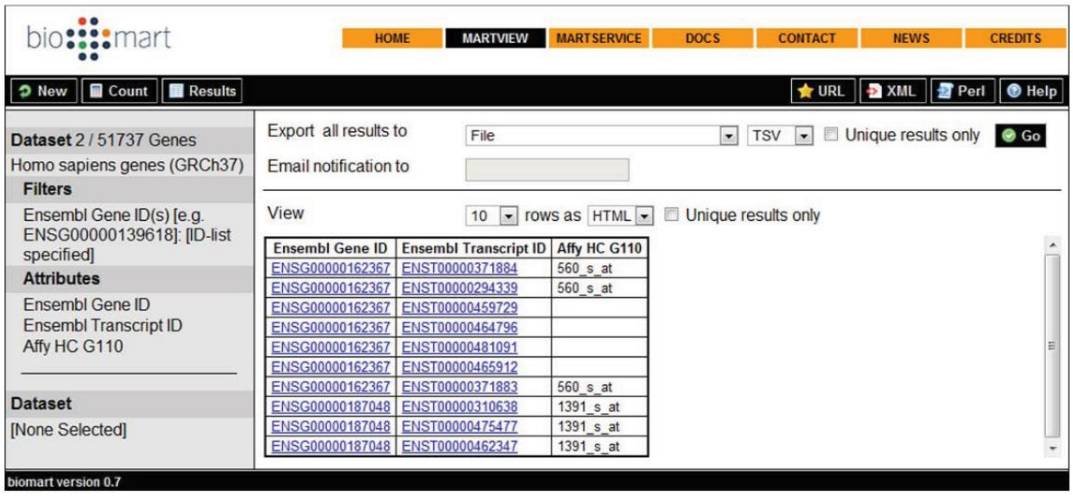

Fig. (2). Biomart example

Originally developed for the Ensemble Genome Browser as EnsMart data warehouse [57], Biomart [58] is a fully generic data integration solution that integrates the most important biological databases (see Fig.2). The LIMBO Architecture [59] is a light-wave approach for molecular database integration that uses a warehouse. The Atlas warehouse [60, 61] integrates data from GenBank, RefSeq, UniPrt, HPRD, BIND, DIP, MINT, IntAct, NCBI Taxonomy, Gene Ontology, OMIM, LocusLink, Entrez Gene and HomoloGene.

The Atlas architecture develops the integration in two levels. The first level uses a common data model to integrate similar types of data and the second level uses ontologies to cross-reference disparate types of data. This architecture provides high-throughput, flexible and complex queries on biological data. Columba [62] integrates databases of annotation of protein structures

like PDB, SWISS-PROT or ENZIME, oriented to structural research. BioWarehouse [63], an open source toolkit for constructing data warehouses, defines a global relational data model for the most important bioinformatics data types. BIOZON [64] integrates heterogeneous data types such as proteins, structures, domain families, protein–protein interactions and cellular pathways. It allows complex queries over interrelated data sources, extending these queries to accommodate fuzzy relationships between proteins and extending the results to sets based on homology instead of direct reference.

The Cancer Research Database [65] uses a mediator to store the data in the warehouse, developed with the aim of finding small molecules that can restore native functions to the p53 cancer mutant genes. E-Fungi [66] is a fungi specific warehouse that integrates a large number of diverse fungal genomes, like the Comparative Fungal Genomics Platform (CFGP) [67], which incorporates fungal genomic data and several analysis tools into a data warehouse. The objective of PharmGKB [68] is to study how genetic variation contributes to drug response, integrating in a warehouse not only pharmacogenomics data but also drug pathways, annotated pharmacogene summaries and relationships among genes, drugs and diseases. In the same line that BioWarehouse, BioDWH [69] aims to provide a java-based kit to construct a biomedical data warehouse using object-relational mapping. GeNS [70] integrates in a warehouse the most representative biological databases (see Fig.3). Finally, bioDBnet [71] is a data warehouse that integrates more than 20 different databases, recognizing more than 100 different types of database types using cross-reference mapping. DW4TR [72] (Data Warehouse for translational Research), is an example of hybrid approach which combines, with a semantic model, a warehouse for the integration of clinical data, that includes clinical data, medical image, genomic analyses, etc., and a federation for the inclusion of external data. Some examples of cheminformatics that follow this approach are Chemspider [73], that integrates molecules of more than 400 chemical data sources. This is an excellent example of "crowdsourcing", because it allows users to contribute in curating the database, in the same way as Wikipedia and similarly to the Distributed Annotation Server [74] philosophy. The use of data sources that can be curtated, both with the crowdsourcing paradigm and for professionals, will improve the integration process. Another solution is Chemistry Connect [75], which was developed by AstraZeneca and integrates several public and private databases into a warehouse with more than 45 million of chemical structures.

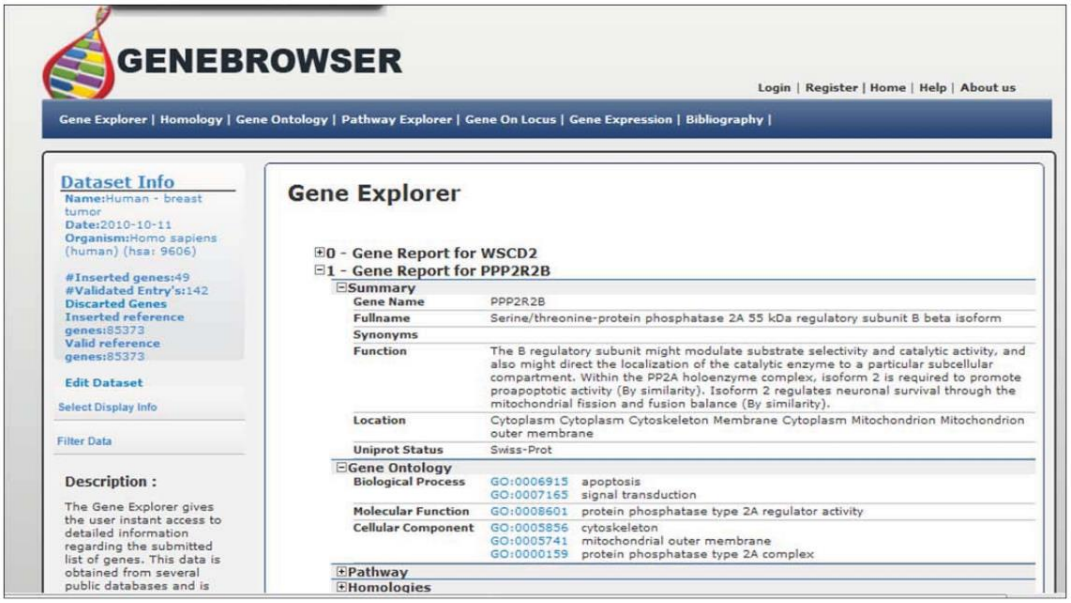

Fig. (3). Gens example.

FEDERATED APPROACHES

Unlike data warehouse, in federated databases, data remains in the original data source. The original databases are autonomous and may be distributed over the net. The federation maintains a common model and the integration is performed mapping data from each source onto this common model using a middleware named wrapper. The federation is shown to the final user as if there was a simple database. This type of integration model solves the updating issue of the data warehouse approach, since the data is always in the original source, so the query always obtains updated data. Biokleisli [76] was a pioneer in the application of this integration model to biological data, in order to answer to questions such as "sample for each gene located in a specific cytogenetic band at a specific human chromosome, as non-human homologues as possible".

The problem of filtering data in federated databases is a complex task, since no data is stored locally. Thus, this has to be done on the fly. Performance is usually also a problem in this model, since it depends on the performance of each data source, to which the integration time must be added. Federated databases use a common model, so it presents the same problems as data warehouse in relation to the representation of different data types.

Database federation is adequate when the researcher requires updated information, or when the researcher must integrate a large amount of data both from private and public databases.

Mediator Schema Approaches

One of the problems of federated databases arises when dealing with several different schemas from different databases. A mediator schema would solve this problem. Federated databases can have relational or semi-structured schemas. One of the first proposals of the mediator schema were The Stanford-IBM Manager of Multiple Information Sources (TSIMMIS) [77] and the SIMS [78].

In general, a mediator schema is a graphical representation of all of the entities and relationships of a specific domain, with the entities as nodes and the relationships as edges. The mediator schemas act as a middleware in a federation of databases, in which data sources are mapped onto the schema of the mediator, defining the entities contained. Queries are performed on the mediator schema, instead of on the common schema of all the databases (see Fig.4). This allows the user to perform more general queries that cannot be answered using a relational database. In addition, it offers the advantage that the mediator schema can be focused on a specific type of user or on a specific set of queries. Given a collection of data sources to integrate, a user can develop a mediator schema focused on his/her data of interest. This would result in a very rich model of a specific subset of the whole data in which he/she is interested without having to develop a global model which has to take into account all the possible queries or data of interest of all the potential users.

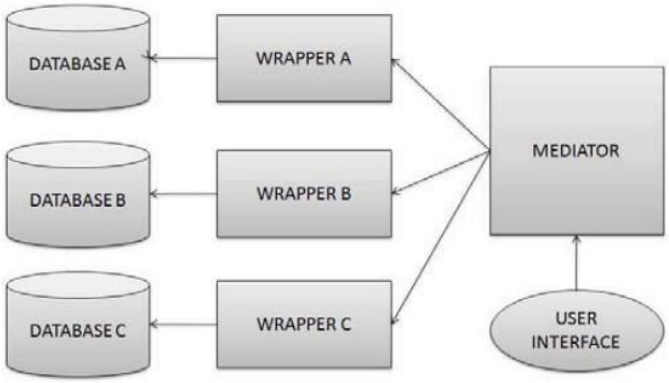

Fig. (4). Federated model with wrapper/mediator.

Mediator schemas can be modular, that is, they can be created and exchanged when necessary. Mediator schemas may be a good selection when researchers need to answer very complex queries from different knowledge domains.

Rather than data integration in the traditional sense where overlapping data elements are transformed into one structure, genomic, transcriptomic and proteomic data need to be linked together using a scaffold that represents their relatedness.

Semantic technologies offer exactly this scaffold. Since genomics provides data on genes, transcriptomic experiments provide data on the transcription of genes (in particular tissues or under specific conditions), and proteomics provides identified peptides, this is not a simple case of transforming different data types and data formats. In this situation there are no common data elements between the data sets.

As we deal here with data relationships which do not involve equality but different degrees of similarity or physical overlap, it is clear that traditional integration methods cannot match these data in a simple manner. However, since these data are mutually related, integration can be achieved by using specific meta-data. Delivering computer understandable meta-data is the basis on which semantic technologies were developed.

Federated Workflows Approaches

Workflow systems have been very used in the last few years to generate complex processes of data processing using web services. These systems have a great capacity of module reuse, allowing a person with little experience in development to visually create complex processing workflows and to automate tasks that previously took weeks. Although there exists certain controversy in relation to naming the workflow systems “federated integration systems”, in this case those that present the following characteristics were considered: 1) a set of services associated to a data source that retrieve its data and transform it according to a common model. These services act like wrappers; 2) a mediator service that divides the queries created by the client to send them to the previously mentioned wrapper services. Subsequently, this service will integrate the data and present it to the client.

Multi-Agent Systems Approaches

Multi-agent systems (MAS) have received considerable attention during recent years. This field has emerged due to the benefits offered by applications or systems that can decide by themselves what they must do in order to satisfy their working principles. An agent is a computational system located in an environment and capable of performing autonomous actions to achieve its design objectives [79]. Woolridge [80] adds to this the following requirements that an agent must meet to be called intelligent: reactivity (i.e., if an agent is capable of perceiving its environment and answering to changes in it), proactivity (the capacity to exhibit behaviors in order to meet objectives having initiative) and social ability (the capacity to interact with other agents). According to this definition, a multi-agent system is a group of agents that collaborate and interact with each other. MAS have certain advantages over isolated agents: reliability, robustness, modularity, scalability, adaptability, concurrence, parallelism and dynamism.

Information integration processes can be considered as complex distributed systems. As postulated in reference [81], multi-agent systems are a great solution to this type of complex distributed systems [82, 83]. In most cases the agent acts as a mediator, normally named mediator agent, representing data following a common model. In addition, a certain number of agents can act as a wrapper. Some agent-based platforms have been developed to carry out information integration tasks, like KRAFT [84].

Federated Examples

In this section we present an overview of all the federated database approaches, including wrapper-only models, wrapper-mediator, agents and semantic integration-based models. BioKleisli [76] [85] was the pioneer in applying the federated approach to biomedical data. The idea of using agents for biomedical data integration problems was proposed by Imai et al.[86], based on the TSMIIS [83] scheme and on the proposal of Bayardo et al.[82] who use agents to solve integration issues. TAMBIS [87, 88] (Transparent Access to Multiple Bioinformatics Information Sources) provides transparent access to several heterogeneous sources. The user queries the biological concept model and this model is mapped onto the source model of each specific data source. DiscoveryLink [89] is a middleware offered by IBM that provides users with a virtual database that allows the user to retrieve data from other databases and offers a query optimizer to retrieve the data. Mork et al.[55] proposes a mediated schema approach to integrate genetic databases. ISYS [90] is a component-based approach that uses a mediated schema and a distributed data model. Karasawas [91] proposes a multi-agent system for data integration, using a wrapper-mediator architecture over a multi-agent system. Information Integration [92] is an extension of the database management system DB2 from IBM, to integrate heterogeneous databases. Semeda (Semantic Metadata Database) [93] is a semantic database integration system that uses a wrapper/mediator schema. BioMediator [94] is a data integration system based on federated databases which uses a mediator schema to answer queries from different data sources. Robinson et al.[95] presents a mediation-based integration system based on the BSML format. The QIS (Query Integrator System) [96, 97] is a database mediator system, that includes OntoMediator, an automatic mediator that uses ontological metadata mappings (see Fig.5) OntoFusion [98] is a mediator architecture that uses Multi Agent Systems to perform the integration. AlzPharm [99] uses RDF (Resource Description Framework) and RDFS (Resource Description Framework Schema) to create a standard data model in order to integrate neuroscience data. SWEDI (Semantic Web Enabled Data Integration) [100] uses ontologies and controlled vocabularies to construct RDF models onto which the data are mapped. Then can run semantic queries over the model. LinkHub uses semantic relations to connect data using a federated approach. It has been tested using Uniprot and East Structural Genomics Consortium. DebugIT [101, 102] is a 7th European Framework Project that aims to improve the detection and elimination of bacteria using IT technologies. One of the scopes of the project was to integrate heterogeneous clinical and biological sources, using a semantic infrastructure. Anwar et al.[103] also applied semantic technologies using RDF to integrate data related with Francisella tularensis novicida. The Prostate Cancer Information System (PCIS) [104] uses ontologies to develop the integration applied to prostate cancer research. The Khaos Ontology-based mediator Framework

[105] is a federated integration architecture that uses a Semantic Directory to register and manage ontologies. This Semantic Directory is used to build a common model represented by ontologies. BioXM [106] uses an object-oriented semantic integration model to build a knowledge management system. This uses semantic networks to link "elements", "relations", "annotations" and "context" concepts with the aim of build the model or can use previously created ontologies. These semantic networks of relationships allow detecting connections, extracting patterns and answering complex questions. To optimize performance, data can be managed alternatively within a relational database or integrated from external data sources, combining the advantages of the data warehouse and the federated approach. The federation of external data sources uses the Biomax software BioRS [107].

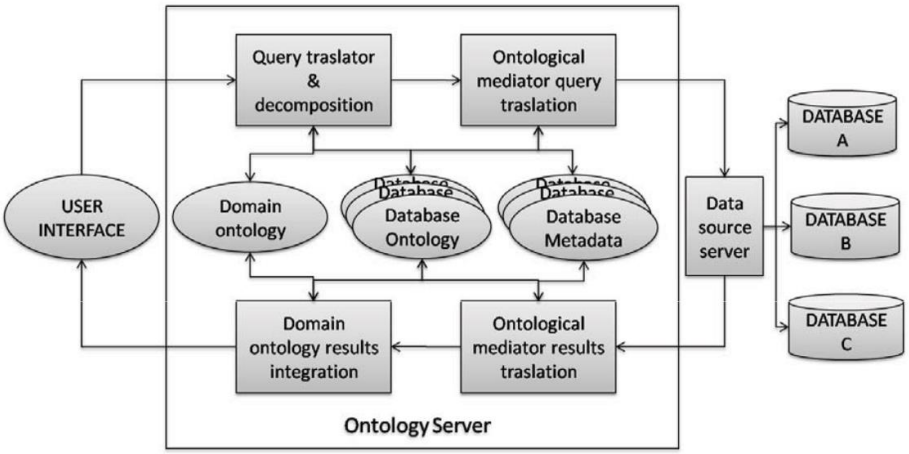

Fig. (5). Ontology Server and QIS architecture.

DISCUSSION

Many of the approaches discussed above are very close to solving the data integration problem in a generic context, or even solve correctly some of the phases part of the integration process. However, none of the analyzed solutions solve completely the biomedical data integration problem in a generic context. Regarding the choice of the most adequate architecture, the link integration approach, although it is one of the most used options to present biomedical databases, is not really considered a type of integration architecture [108]. This is because it does not provide enough functionality regarding the queries and it does not adapt well to great amounts of data and the changing nature both of the data and the interfaces of the data sources. Furthermore, it is very vulnerable to name conflicts and ambiguities. The federated approach presents several advantages over warehouse, such as not requiring a high investment on hardware. However, the federated approach always provides the last version of the data, since the data is always retrieved from the original data source. In contrast, warehouses must be updated frequently in order to provide the most recent data. The main advantages of the warehouses lie in their centralization, allowing an easier scalability, better performance in queries, great availability and greater control over the data. On the contrary, federated approaches present a more complicated solution, more latency in retrieving the data and loss of control over these as they belong to the original data source. In addition, the availability of the whole system depends on the availability of all the data sources. Regarding the data models, the federated approach shows a clear advantage over the warehouse approach: since the data models of the latter were created ad-hoc, this generally implies that part of this data and its relationships, which are present in the original data sources, have been lost. In the federated approach, however, access mechanisms are usually more sophisticated, allowing directly querying the data bases (as long as the interface allows it), thus making possible to access not only the data but also the existing relationships.

CONCLUSIONS AND FUTURE TRENDS

In this review we analyze most of the existing approaches to integrate biomedical data. These solutions can be used in a lot of contexts, such as drug discovery, clinical issues, system biology, etc. As far as we are aware of, no current integration solution addresses completely the overlapping nature of integrated data in a generic context. The majority of the existing solutions achieve horizontal integration; data sources are treated as complementary to one another, and issues associated with data aggregation are ignored.

After analyzing the pros and cons of the different types of architecture, we could assert that the federated approach, in general, provides a better solution to the biomedical data integration problem, as long as the integration problem is not a monolithic application, the requisites are known in advance, access to the whole database is possible and the data must not be completely up to date. On the contrary, this approach is not recommended when a total control over the data and great performance are required.

As said in reference [64], "Data integrity, consistency, redundancy, connectivity, updatability, expandability and complex and 'fuzzy' queries are the problems associated with data integration, which arise from the nature of heterogeneous data and the lack of unified ontology". Therefore, there is a need for integration systems that are able to recognize different ontologies and semantics of the data. Yet integration systems should also provide an environment that allows users to integrate their own data and customize the system. After analyzing this solution, we noted that there is a clear trend towards the use of semantic technologies. The semantic annotation, mapping and querying process is increasingly being used and offers a very suitable approach in data integration. Ontologies and other semantic technologies are a great tool; however, their quality should be improved, as well as their correct use in the data source annotation process. In Goble's work [13], in order to defend the idea of light integration using mashups, it is argued that, for better or worse, in bioinformatics, application development is based on the "just in time, just enough" mantra. Moreover, it is also said that biology is what really matters, not engineering, because this last one presents too complex solutions that take too long to be developed and are not adequate to the user's needs. After analyzing the existing approaches on biomedical data integration, it can be concluded that instead of basing the development of integration systems on a specific architecture or model, developing a methodology for biomedical data integration, which allows providing dimensioned, correct and adapted solutions to the problem's needs, would be more adequate.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Vanessa Aguiar-Pulido and Cristian R. Munteanu acknowledge the funding support for a research position by the "Plan I2C" and an "Isidro Parga Pondal" Program both from Xunta de Galicia, Spain (supported by the European Social Fund). This work is supported by the following projects: "Galician Network for Colorectal Cancer Research" (REGICC, Ref. 2009/58) from the General Directorate of Research, Development and Innovation of Xunta de Galicia, "Ibero-American Network of the Nano-Bio-Info-Cogno Convergent Technologies", Ibero-NBIC Network (209RT-0366) funded by CYTED (Spain), grant Ref. PIO52048 and RD07/0067/0005 funded by the Carlos III Health Institute and "PHR2.0: Registro Personal de Salud en Web 2.0" (Ref. TSI-020110-2009-53) funded by the Spanish Ministry of Industry, Tourism and Trade.

REFERENCES

[1] Williams, N. How to get databases talking the same language. Science,1997,275, 301-302.
[2] Sujansky, W. Heterogeneous database integration in biomedicine. J. Biomed. Inform.,2001,34, 285-298.
[3] Antezana, E.; Kuiper, M.; Mironov, V. Biological knowledge management: the emerging role of the Semantic Web technologies. Brief. Bioinform.,2009,10, 392-407.
[4] Wong, L. Technologies for integrating biological data. Brief. Bioinform.,2002,3, 389-404.
[5] Stein, L.D. Integrating biological databases. Nat. Rev. Genet.,2003,4, 337-345.
[6] Köhler, J. Integration of life science databases. Drug Discov. Today: BIOSILICO,2004,2, 9.
[7] Philippi, S.; Kohler, J. Addressing the problems with life-science databases for traditional uses and systems biology. Nat. Rev. Genet.,2006,7, 482-488.
[8] Brazhnik, O.; Jones, J.F. Anatomy of data integration. J. Biomed. Inform.,2007,40, 252-269.
[9] Louie, B.; Mork, P.; Martin-Sanchez, F.; Halevy, A.; Tarczy-Hornoch, P. Data integration and genomic medicine. J. Biomed. Inform.,2007,40, 5-16.
[10] Burgun, A.; Bodenreider, O. Accessing and integrating data and knowledge for biomedical research. Yearb. Med. Inform.,2008, 91-101.
[11] Cheung, K.-H.; Kashyap, V.; Luciano, J.S.; Chen, H.; Wang, Y.; Stephens, S. Semantic mashup of biomedical data. J. Biomed. Inform.,2008,41, 683-686.
[12] Cheung, K.H.; Yip, K.Y.; Townsend, J.P.; Scotch, M. HCLS 2.0/3.0: health care and life sciences data mashup using Web 2.0/3.0. J. Biomed. Inform.,2008,41, 694-705.
[13] Goble, C.; Stevens, R. State of the nation in data integration for bioinformatics. J. Biomed. Inform.,2008,41,687-693.
[14] Philippi, S. Data and knowledge integration in the life sciences. Brief. Bioinform.,2008,9, 451.
[15] Akula, S.P.; Miriyala, R.N.; Thota, H.; Rao, A.A.; Gedela, S. Techniques for integrating -omics data. Bioinformation,2009,3,284-286.
[16] Ruttenberg, A.; Clark, T.; Bug, W.; Samwald, M.; Bodenreider, O.; Chen, H.; Doherty, D.; Forsberg, K.; Gao, Y.; Kashyap, V.; Kinoshita, J.; Luciano, J.; Marshall, M.S.; Ogbuji, C.; Rees, J.; Stephens, S.; Wong, G.T.; Wu, E.; Zaccagnini, D.; Hongsermeier, T.; Neumann, E.; Herman, I.; Cheung, K.H. Advancing translational research with the Semantic Web. BMC Bioinformatics,2007,8, S2.
[17] Zhang, Z.; Cheung, K.H.; Townsend, J.P. Bringing Web 2.0 to bioinformatics. Brief. Bioinform.,2009,10, 1-10.
[18] Cheung, K.H.; Frost, H.R.; Marshall, M.S.; Prud'hommeaux, E.; Samwald, M.; Zhao, J.; Paschke, A. A journey to Semantic Web query federation in the life sciences. BMC Bioinformatics,2009,10, S10.
[19] Lambrix, P.; Strömbäck, L.; Tan, H. Information Integration in Bioinformatics with Ontologies and Standards. In: Bry F, Maluszynski J, ed.^eds., Semantic Techniques for the Web. Springer Berlin / Heidelberg, 2009; pp. 343-376.
[20] Martin-Sanchez, F.; Iakovidis, I.; Norager, S.; Maojo, V.; de Groen, P.; Van der Lei, J.; Jones, T.; Abraham-Fuchs, K.; Apweiler, R.; Babic, A.; Baud, R.; Breton, V.; Cinquin, P.; Doupi, P.; Dugas, M.; Eils, R.; Engelbrecht, R.; Ghazal, P.; Jehenson, P.; Kulikowski, C.; Lampe, K.; De Moor, G.; Orphanoudakis, S.; Rossing, N.; Sarachan, B.; Sousa, A.; Spekowius, G.; Thireos, G.; Zahlmann, G.; Zvarova, J.; Hermosilla, I.; Vicente, F.J. Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. J. Biomed. Inform.,2004,37, 30-42.
[21] Sax, U.; Schmidt, S. Integration of genomic data in Electronic Health Records--opportunities and dilemmas. Methods Inf. Med.,2005,44, 546-550.
[22] Mitchell, D.R.; Mitchell, J.A. Status of clinical gene sequencing data reporting and associated risks for information loss. J. Biomed. Inform.,2007,40, 47-54.
[23] Adida, B.; Kohane, I.S. GenePING: secure, scalable management of personal genomic data. BMC Genomics,2006,7, 93.
[24] Brazas, M.D.; Yamada, J.T.; Ouellette, B.F. Providing web servers and training in bioinformatics: 2010 update on the bioinformatics links directory. Nucleic Acids Res.,2010,38, W3-6.
[25] Blake, J.A.; Bult, C.J. Beyond the data deluge: Data integration and bio-ontologies. J. Biomed. Inform.,2006,39, 314-320.
[26] Bodenreider, O.; Stevens, R. Bio-ontologies: current trends and future directions. Brief. Bioinform.,2006,7, 256-274.
[27] Rubin, D.L.; Shah, N.H.; Noy, N.F. Biomedical ontologies: a functional perspective. Brief. Bioinform.,2008,9, 75-90.
[28] Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L.J.; Eilbeck, K.; Ireland, A.; Mungall, C.J.; Leontis, N.; Rocca-Serra, P.; Ruttenberg, A.; Sansone, S.A.; Scheuermann, R.H.; Shah, N.; Whetzel, P.L.; Lewis, S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol.,2007,25, 1251-5.
[29] Dumontier, M.; Villanueva-Rosales, N. Towards pharmacogenomics knowledge discovery with the semantic web. Brief. Bioinform.,2009,10, 153-163.
[30] Feigenbaum, L.; Herman, I.; Hongsermeier, T.; Neumann, E.; Stephens, S. The semantic web in action. Scientific American,2007,297, 90-97.

[31] Neumann, E. A life science Semantic Web: are we there yet? Sci. STKE.,2005,2005, 22.
[32] Friedman, C.; Borlawsky, T.; Shagina, L.; Xing, H.R.; Lussier, Y.A. Bio-Ontology and text: bridging the modeling gap. Bioinformatics,2006,22, 2421-2429.
[33] Good, B.M.; Wilkinson, M.D. The life sciences semantic web is full of creeps! Brief. Bioinform.,2006,7,275-286.
[34] Chen, H.; Ding, L.; Wu, Z.; Yu, T.; Dhanapalan, L.; Chen, J.Y. Semantic web for integrated network analysis in biomedicine. Brief. Bioinform.,2009,10, 177-192.
[35] Cheung, K.-H.; Prud'hommeaux, E.; Wang, Y.; Stephens, S. Semantic web for health care and life sciences: a review of the state of the art. Brief. Bioinform.,2009,10,111-113.
[36] Pasquier, C. Biological data integration using semantic web technologies. Biochimie.,2008,90, 584-594.
[37] Etzold, T.; Ulyanov, A.; Argos, P. SRS: information retrieval system for molecular biology data banks. Methods Enzymol.,1996,266, 114-128.
[38] Zdobnov, E.M.; Lopez, R.; Apweiler, R.; Etzold, T. The EBI SRS server--recent developments. Bioinformatics,2002,18, 368-373.
[39] Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J.P. ChEMBL: A large-scale bioactivity database for drug discovery. Nucleic Acids Res.,2012,40, D1100-1107.
[40] Baxevanis, A.D. Searching NCBI databases using Entrez. Curr. Protoc. Bioinformatics,2008, Chapter 1, Unit 1 3.
[41] Li, Q.; Cheng, T.; Wang, Y.; Bryant, S.H. PubChem as a public resource for drug discovery. Drug Discov. Today,2010,15, 1052-1057.
[42] Kersey, P.; Bower, L.; Morris, L.; Horne, A.; Petryszak, R.; Kanz, C.; Kanapin, A.; Das, U.; Michoud, K.; Phan, I.; Gattiker, A.; Kulikova, T.; Faruque, N.; Duggan, K.; McLaren, P.; Reimholz, B.; Duret, L.; Penel, S.; Reuter, I.; Apweiler, R. Integr8 and Genome Reviews: integrated views of complete genomes and proteomes. Nucleic Acids Res.,2005,33, D297-302.
[43] Oliveira, J.L.; Dias, G.; Oliveira, I.; Rocha, P.; Hermosilla, I.; Vicente, J.; Spiteri, I.; Martin-Sánchez, F.; Pereira, A.S. DiseaseCard: A web-based tool for the collaborative integration of genetic and medical information. In: Barreiro JM, Martin-Sanchez F, Maojo V, Sanz F, ed.^eds., Biological and medical data analysis. Springer Berlin / Heidelberg, 2004, pp. 409-417.
[44] Chaudhuri, S.; Dayal, U. An overview of data warehousing and OLAP technology. SIGMOD Rec.,1997,26, 65-74.
[45] Schönbach, C.; Kowalski-Saunders, P.; Brusic, V. Data warehousing in molecular biology.Brief. Bioinform.,2000,1,190-198.
[46] Goble, C.; Stevens, R.; Hull, D.; Wolstencroft, K.; Lopez, R. Data curation + process curation=data integration + science. Brief. Bioinform.,2008,9, 506-517.
[47] Ritter, O.; Kocab, P.; Senger, M.; Wolf, D.; Suhai, S. Prototype implementation of the integrated genomic database. Comput. Biomed. Res.,1994,27, 97-115.
[48] Nadkarni, P.M.; Brandt, C.; Frawley, S.; Sayward, F.G.; Einbinder, R.; Zelterman, D.; Schacter, L.; Miller, P.L. Managing attribute--value clinical trials data using the ACT/DB client-server database system. J. Am. Med. Inform. Assoc.,1998,5, 139-51.
[49] Critchlow, T.; Fidelis, K.; Ganesh, M.; Musick, R.; Slezak, T. DataFoundry: information management for scientific data. IEEE. Trans. Inf. Technol. Biomed.,2000,4, 52-7.
[50] Cornell, M.; Paton, N.W.; Shengli, W.; Goble, C.A.; Miller, C.J.; Kirby, P.; Eilbeck, K.; Brass, A.; Hayes, A.; Oliver, S.G. GIMS-a data warehouse for storage and analysis of genome sequence and functional data. In: ed.^eds., Bioinformatics and Bioengineering Conference, 2001. Proceedings of the IEEE 2nd International Symposium on, 2001;pp. 15-22.
[51] Bukhman, Y.V.; Skolnick, J. BioMolQuest: integrated database-based retrieval of protein structural and functional information. Bioinformatics,2001,17, 468-478.
[52] Eckman, B.A.; Kosky, A.S.; Laroco, J., Leonardo A. Extending traditional query-based integration approaches for functional characterization of post-genomic data. Bioinformatics,2001,17, 587-601.
[53] Hubbard, T.; Barker, D.; Birney, E.; Cameron, G.; Chen, Y.; Clark, L.; Cox, T.; Cuff, J.; Curwen, V.; Down, T.; Durbin, R.; Eyras, E.; Gilbert, J.; Hammond, M.; Huminiecki, L.; Kasprzyk, A.; Lehvaslaiho, H.; Lijnzaad, P.; Melsopp, C.; Mongin, E.; Pettett, R.; Pocock, M.; Potter, S.; Rust, A.; Schmidt, E.; Searle, S.; Slater, G.; Smith, J.; Spooner, W.; Stabenau, A.; Stalker, J.; Stupka, E.; Ureta-Vidal, A.; Vastrik, I.; Clamp, M. The Ensembl genome database project. Nucleic Acids Res.,2002,30, 38-41.
[54] Haas, L.M.; Schwarz, P.M.; Kodali, P.; Kotlar, E.; Rice, J.E.; Swope, W.C. DiscoveryLink: a system for integrated access to life sciences data sources. IBM Syst. J.,2001,40, 489-511.
[55] Mork, P.; Halevy, A.; Tarczy-Hornoch, P. A model for data integration systems of biomedical data applied to online genetic databases. Proc. AMIA. Symp.,2001, 473-7.
[56] Karolchik, D.; Baertsch, R.; Diekhans, M.; Furey, T.S.; Hinrichs, A.; Lu, Y.T.; Roskin, K.M.; Schwartz, M.; Sugnet, C.W.; Thomas, D.J.; Weber, R.J.; Haussler, D.; Kent, W.J. The UCSC Genome Browser Database. Nucleic Acids Res.,2003,31, 51-4.

[57] Kasprzyk, A.; Keefe, D.; Smedley, D.; London, D.; Spooner, W.; Melsopp, C.; Hammond, M.; Rocca-Serra, P.; Cox, T.; Birney, E. EnsMart: a generic system for fast and flexible access to biological data. Genome Res.,2004,14, 160-9.
[58] Smedley, D.; Haider, S.; Ballester, B.; Holland, R.; London, D.; Thorisson, G.; Kasprzyk, A. BioMart--biological queries made easy. BMC Genomics,2009,10, 22.
[59] Philippi, S. Light-weight integration of molecular biological databases. Bioinformatics,2004,20, 51-57.
[60] Birch, P.; Friedman, J.M. Utility and limitations of genetic disease databases in clinical genetics research: a neurofibromatosis 1 database example. Am. J. Med. Genet. C. Semin. Med. Genet.,2004,125C, 42-9.
[61] Shah, S.P.; Huang, Y.; Xu, T.; Yuen, M.M.; Ling, J.; Ouellette, B.F. Atlas - a data warehouse for integrative bioinformatics. BMC Bioinformatics,2005,6, 34.
[62] Trissl, S.; Rother, K.; Muller, H.; Steinke, T.; Koch, I.; Preissner, R.; Frommel, C.; Leser, U. Columba: an integrated database of proteins, structures, and annotations. BMC Bioinformatics,2005,6, 81.
[63] Lee, T.J.; Pouliot, Y.; Wagner, V.; Gupta, P.; Stringer-Calvert, D.W.; Tenenbaum, J.D.; Karp, P.D. BioWarehouse: a bioinformatics database warehouse toolkit. BMC Bioinformatics,2006,7, 170.
[64] Birkland, A.; Yona, G. BIOZON: a hub of heterogeneous biological data. Nucleic Acids Res.,2006,34, D235-242.
[65] Bichutskiy, V.Y.; Colman, R.; Brachmann, R.K.; Lathrop, R.H. Heterogeneous biomedical database integration using a hybrid strategy: a p53 cantcer research database. Cancer Inform.,2007,2, 277-287.
[66] Hedeler, C.; Wong, H.M.; Cornell, M.J.; Alam, I.; Soanes, D.M.; Rattray, M.; Hubbard, S.J.; Talbot, N.J.; Oliver, S.G.; Paton, N.W. e-Fungi: a data resource for comparative analysis of fungal genomes. BMC Genomics,2007,8, 426.
[67] Park, J.; Park, B.; Jung, K.; Jang, S.; Yu, K.; Choi, J.; Kong, S.; Kim, S.; Kim, H.; Kim, J.F.; Blair, J.E.; Lee, K.; Kang, S.; Lee, Y.H. CFGP: a web-based, comparative fungal genomics platform. Nucleic Acids Res.,2008,36, D562-571.
[68] Gong, L.; Owen, R.P.; Gor, W.; Altman, R.B.; Klein, T.E. PharmGKB: an integrated resource of pharmacogenomic data and knowledge. Curr Protoc Bioinformatics,2008; Chapter 14: Unit1417.
[69] Topel, T.; Kormeier, B.; Klassen, A.; Hofestadt, R. BioDWH: a data warehouse kit for life science data integration. J. Integr. Bioinform.,2008, 5.
[70] Arrais, J.; Pereira, J.; Fernandes, J.; Oliveira, J. GeNS: A biological data integration platform. In: ed.^eds., 2009, pp. 850-855.
[71] Mudunuri, U.; Che, A.; Yi, M.; Stephens, R.M. bioDBnet: The biological database network. Bioinformatics,2009,25, 555-556.
[72] Hu, H.; Correll, M.; Kvecher, L.; Osmond, M.; Clark, J.; Bekhash, A.; Schwab, G.; Gao, D.; Gao, J.; Kubatin, V.; Shriver, C.D.; Hooke, J.A.; Maxwell, L.G.; Kovatich, A.J.; Sheldon, J.G.; Liebman, M.N.; Mural, R.J. DW4TR: A data warehouse for translational research. J. Biomed. Inform.,2011,44, 1004-19.
[73] Hettne, K.M.; Williams, A.J.; van Mulligen, E.M.; Kleinjans, J.; Tkachenko, V.; Kors, J.A. Automatic vsmanual curation of a multi-source chemical dictionary: the impact on text mining. J. Cheminform.,2010,2, 3.
[74] Dowell, R.D.; Jokerst, R.M.; Day, A.; Eddy, S.R.; Stein, L. The distributed annotation system. BMC Bioinformatics,2001,2, 7.
[75] Muresan, S.; Petrov, P.; Southan, C.; Kjellberg, M.J.; Kogej, T.; Tyrchan, C.; Varkonyi, P.; Xie, P.H. Making every SAR point count: the development of Chemistry Connect for the large-scale integration of structure and bioactivity data. Drug Discov. Today,2011,16, 1019-1030.
[76] Davidson, S.B.; Overton, C.; Tannen, V.; Wong, L. BioKleisli: a digital library for biomedical researchers. Int. J. Digital Libraries,1997,1, 36-53.
[77] Chawathe, S.; Garcia-Molina, H.; Hammer, J.; Ireland, K.; Papakonstantinou, Y.; Ullman, J.; Widom, J. The TSIMMIS Project: Integration of Heterogenous Information Sources. In: ed.^eds., Information Processing Society of Japan (IPSJ 1994): Tokyo, Japan, 1994.
[78] Arens, Y.; Hsu, C.-N.; Knoblock, C.A. Query processing in the SIMS information mediator. In: ed.^eds., Readings in agents. Morgan Kaufmann Publishers Inc., 1998; pp. 82-90.
[79] Wooldridge, M. An introduction to multiagent systems. John Wiley & Sons LTD. 2002.
[80] Wooldridge, M.; R, J. Intelligent Agents: theory and practice. Knowl. Eng. Rev.,1995,10, 38.
[81] Jennings, N.R. An agent-based approach for building complex software systems. Commun. ACM.,2001,44, 35-41.
[82] Bayardo, R.J. Jr.; Bohrer, W.; Brice, R.; Cichocki, A.; Fowler, J.; Helal, A.; Kashyap, V.; Ksiezyk, T.; Martin, G.; Nodine, M.; Rashid, M.; Rusinkiewicz, M.; Shea, R.; Unnikrishnan, C.; Unruh, A.; Woelk, D. InfoSleuth: agent-based semantic integration of information in open and dynamic environments. SIGMOD Rec.,1997,26,195-206.
[83] Garcia-Molina, H.; Papakonstantinou, Y.; Quass, D.; Rajaraman, A.; Sagiv, Y.; Ullman, J.; Vassalos, V.; Widom, J. The TSIMMIS Approach to Mediation: Data Models and Languages. J. Intell. Inf. Syst.,1997,8, 117-132.
[84] Preece, A.; Hui, K.; Gray, A.; Marti, P.; Bench-Capon, T.; Jones, D.; Cui, Z. The KRAFT architecture for knowledge fusion and transformation. Knowledge-Based Systems,2000,13, 113-120.

[85] Stein, L.D.; Thierry-Mieg, J. Scriptable access to the Caenorhabditis elegans genome sequence and other ACEDB databases. Genome Res.,1998,8, 1308-1315.
[86] Imai, T.; Matsuda, H.; Sekihara, T.; Nakanishi, M.; Hashimoto, A. Implementing an integrated system for heterogeneous molecular biology databases with intelligent agents. In: ed.^eds., Communications, Computers and Signal Processing, 1997. '10 Years PACRIM 1987-1997 - Networking the Pacific Rim'. 1997 IEEE Pacific Rim Conference on, 1997; pp. 807-810 vol.2.
[87] Baker, P.G.; Brass, A.; Bechhofer, S.; Goble, C.; Paton, N.; Stevens, R. TAMBIS-transparent access to multiple bioinformatics information sources. Proc. Int. Conf. Intell. Syst. Mol. Biol.,1998,6, 25-34.
[88] Stevens, R.; Baker, P.; Bechhofer, S.; Ng, G.; Jacoby, A.; Paton, N.W.; Goble, C.A.; Brass, A. TAMBIS: transparent access to multiple bioinformatics information sources. Bioinformatics,2000,16, 184-5.
[89] Haas, L.M.; Schwarz, P.M.; Kodali, P.; Kotlar, E.; Rice, J.E.; Swope, W.C. DiscoveryLink: A system for integrated access to life sciences data sources. Ibm Syst. J.,2001,40, 489-511.
[90] Siepel, A.; Farmer, A.; Tolopko, A.; Zhuang, M.; Mendes, P.; Beavis, W.; Sobral, B. ISYS: a decentralized, component-based approach to the integration of heterogeneous bioinformatics resources. Bioinformatics,2001,17, 83-94.
[91] Karasavvas, K.; Burger, A.; Baldock, R. A Multi-agent Bioinformatics Integration System with Adjustable Autonomy. In: Ishizuka M, Sattar A, ed.^eds., PRICAI 2002: Trends in Artificial Intelligence. Springer Berlin / Heidelberg, 2002; pp. 171-187.
[92] Arenson, A.D. Federating data with Information Integrator. Brief. Bioinform.,2003,4, 375-381.
[93] Köhler, J.; Philippi, S.; Lange, M. SEMEDA: ontology based semantic integration of biological databases. Bioinformatics,2003,19, 2420-2427.
[94] Donelson, L.; Tarczy-Hornoch, P.; Mork, P.; Dolan, C.; Mitchell, J.A.; Barrier, M.; Mei, H. The BioMediator system as a data integration tool to answer diverse biologic queries. Stud. Health Technol. Inform.,2004,107,768-772.
[95] Robinson, A.; Rahayu, W. Genome Database Integration. In: Laganà A, Gavrilova ML, Kumar V, Mun Y, Tan CJK, Gervasi O, ed.^eds., Computational Science and Its Applications – ICCSA 2004. Springer Berlin / Heidelberg, 2004; pp. 443-453.
[96] Marenco, L.; Wang, T.Y.; Shepherd, G.; Miller, P.L.; Nadkarni, P. QIS: A framework for biomedical database federation. J. Am. Med. Inform. Assoc.,2004,11, 523-534.
[97] Marenco, L.; Wang, R.; Nadkarni, P. automated database mediation using ontological metadata mappings. J. Am. Med. Inform. Assoc., 2009,16, 723-737.
[98] Alonso-Calvo, R.; Maojo, V.; Billhardt, H.; Martin-Sanchez, F.; García-Remesal, M.; Pérez-Rey, D. An agent- and ontology-based system for integrating public gene, protein, and disease databases. J. Biomed. Inform.,2007,40, 17-29.
[99] Lam, H.Y.; Marenco, L.; Clark, T.; Gao, Y.; Kinoshita, J.; Shepherd, G.; Miller, P.; Wu, E.; Wong, G.T.; Liu, N.; Crasto, C.; Morse, T.; Stephens, S.; Cheung, K.H. AlzPharm: integration of neurodegeneration data using RDF. BMC Bioinformatics,2007,8, S4.
[100] Post, L.J.G.; Roos, M.; Marshall, M.S.; van Driel, R.; Breit, T.M. A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data. Bioinformatics,2007,23, 3080-3087.
[101] Lovis, C.; Colaert, D.; Stroetmann, V.N. DebugIT for patient safety - improving the treatment with antibiotics through multimedia data mining of heterogeneous clinical data. Stud. Health Technol. Inform.,2008,136, 641-6.
[102] Teodoro, D.; Choquet, R.; Pasche, E.; Gobeill, J.; Daniel, C.; Ruch, P.; Lovis, C. Biomedical data management: a proposal framework. Stud. Health Technol. Inform.,2009,150, 175-9.
[103] Anwar, N.; Hunt, E. Francisella tularensis novicida proteomic and transcriptomic data integration and annotation based on semantic web technologies. BMC Bioinformatics,2009,10, S3.
[104] Min, H.; Manion, F.J.; Goralczyk, E.; Wong, Y.-N.; Ross, E.; Beck, J.R. Integration of prostate cancer clinical data using an ontology. J. Biomed. Inform.,2009,42, 1035-1045.
[105] Roldan-Garcia Mdel, M.; Navas-Delgado, I.; Kerzazi, A.; Chniber, O.; Molina-Castro, J.; Aldana-Montes, J.F. KA-SB: from data integration to large scale reasoning. BMC Bioinformatics,2009,10: S5.
[106] Maier, D.; Kalus, W.; Wolff, M.; Kalko, S.G.; Roca, J.; Marin de Mas, I.; Turan, N.; Cascante, M.; Falciani, F.; Hernandez, M.; Villa-Freixa, J.; Losko, S. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC Syst. Biol.,2011,5, 38.
[107] Kaps, A.; Dyshlevoi, K.; Heumann, K.; Jost, R.; Kontodinas, I.; Wolff, M.; Hani, J. The BioRS(TM) Integration and Retrieval System: An open system for distributed data integration. J. Integr. Bioinform.,2006,3, 44.
[108] Davidson, S.B.; Overton, C.; Buneman, P. Challenges in integrating biological data sources. J. Comput. Biol.,1995,2, 557-572