frontiers in
PHYSIOLOGY

REVIEW ARTICLE
published: 03 December 2010
doi: 10.3389/fphys.2010.00147

Current trends and new challenges of databases and web applications for systems driven biological research

Pradeep Kumar Sreenivasaiah and Do Han Kim*

Systems Biology Research Center and College of Life Science, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea

Edited by:
Taishin Nomura, Osaka University, Japan

Reviewed by:
Hiroaki Wagatsuma, Kyushu Institute of Technology, Japan
Hideo Matsuda, Osaka University, Japan

****Correspondence:***
Do Han Kim, Systems Biology Research Center, Gwangju Institute of Science and Technology, 1 Oryong-dong, Buk-gu, Gwangju 500-712, Korea.
e-mail: dhkim@gist.ac.kr

Dynamic and rapidly evolving nature of systems driven research imposes special requirements on the technology, approach, design and architecture of computational infrastructure including database and Web application. Several solutions have been proposed to meet the expectations and novel methods have been developed to address the persisting problems of data integration. It is important for researchers to understand different technologies and approaches. Having familiarized with the pros and cons of the existing technologies, researchers can exploit its capabilities to the maximum potential for integrating data. In this review we discuss the architecture, design and key technologies underlying some of the prominent databases and Web applications. We will mention their roles in integration of biological data and investigate some of the emerging design concepts and computational technologies that are likely to have a key role in the future of systems driven biomedical research.

Keywords: data integration, data warehouse, federated database, web service, semantic web, ontology, web2.0, grid computing

INTRODUCTION

Nobel Laurate Ivan P. Pavlov tried to understand basic animal physiology by methodically planning surgical experiments, which he believed could advance knowledge in humans. His main contributions were in studying neuronal input to the stomach and pancreas triggering secretions of acid and digestive enzymes with the anticipation of the ingestion of a desirable food. This cephalic phase experiment by Pavlov was brilliant demonstration of systems approach studying interaction of multiple subsystems like brain and gut, even though investigating techniques used by Pavlov are now considered as conventional approach. The Nobel Prize, awarded to Pavlov was the first ever awarded for the studies in integrated systems physiology (Wood, 2004).

We want to emphasize that the nature of systems physiology, from the beginning, has been interdisciplinary. Systems physiology promotes the sense that biological components are not merely an isolated entity, but, on the contrary, is part of highly interconnected coherently functioning dynamic network. The field particularly concerns with recognizing the importance of interactions between biological components and the consequences of those interactions. Thus systems physiology embody holistic views as to how molecules, pathways and networks interact to establish a functioning system at different levels of organization from molecules, organelles, cells, tissues to organs, and even to entire organisms, and further and how malformations in these system leads to diseases? So with all the knowledge in hand we can step forward to develop detailed computational model of human body.

Progress in systems driven research [e.g., systems biology, physiome, systems physiology systems pharmacology, virtual physiological human (VPH), personal health systems, life science e-infrastructures] is significantly driven by development of suitable computational infrastructure including tools and information resources. Over the past few years a variety of high- throughput methodologies (~omics) were developed that has enabled large-scale studies of the biological components at different organizational levels and various scales: genome; interactome; cellular function; tissue and whole-organ structure–function relationship; and integrative functions of the whole organism (e.g behavior and consciousness) to name the few. This has generated massive amounts of data about biological components in multiple sets of experimental conditions. Mostly the contributions are from laboratories around the world following proprietary standards, techniques and methods. Systems biologists seek to integrate and interpret such massive amounts of highly heterogeneous information to understand how biological system functions. Issues and challenges in data integration problems has been meticulously addressed by development of several biological data standardization initiatives (e.g., SBML, Finney and Hucka, 2003; insilicoML, Asai et al., 2008; and CellML, Lloyd et al., 2004), ontologies (e.g., GO, Ashburner et al., 2000; SBO, Le Novere, 2006; and BioPAX, 2006), and establishing large software infrastructures and tools (e.g., NCBI, Sayers et al., 2010; EBI, Brooksbank et al., 2005; Bioconductor, Gentleman et al., 2004; eScience, Hey and Trefethen, 2003). This progress in the field of biomedical data integration has resulted in development of a good methodological and technological framework and much of this has happened just in last decade. But like others, we also believe that cornucopia of the best practices is still evolving and major developments in data integration are underway. The anticipation is, once established such computational infrastructure will enable collaborative investigation of complex biological systems and will help to tackle challenges underlying systems physiology research.

The objective of this review paper will be to understand the major data integration challenges. We will discuss about the kinds of integration approaches and technologies that have been tried to meet the challenges. The progress, pros and cons of the major technologies supporting systems research will be reflected in the substance of our

discussions. We will also discuss the future technologies and new challenges that are anticipated and might help progress of systems physiology. We have compiled a glossary of terms and list of useful databases and Web applications and made it available as supplementary online web pages, which is accessible at this Web address: http://cidms.org/systems_research/resource.html.

OVERVIEW OF PROBLEMS IN DATABASES AND WEB APPLICATIONS IN INTEGRATING INFORMATION FOR SYSTEMS SCALE ANALYSIS

Systems biological research is dynamically evolving in a rapid pace. Research in systems biology is becoming more sophisticated in terms of the capabilities expected from the databases and Web applications. Attributed to the very complex nature of the biological systems, enumerating every requirement systematically, for developing computing infrastructure, is highly difficult. To design and build such computational systems involves usage of numerous standards, technologies, frameworks and tool kits which are complex, increasingly expensive to build and maintain, and requires meticulous planning and management.

In our view those critical issues associated with databases and Web applications supporting systems research are listed in BOX1. Current emerging data integration approaches and technologies should address these issues in order to facilitate continued progress in systems research.

BOX 1 | A summary of critical issues

Data availability: Data availability deals with the issues associated with accessing data in public and private setup which by large is influenced by the institutional policy differences.

Data quantity: Systems research is an iterative and data intensive. Current data will give rise to new information and models and inturn will result in more new data with variations. This cycle continues, and data volume increases exponentially. Therefore management of data quantity is crucial for systems driven research.

Data quality: Data quality describes a set of data properties describing their ability to satisfy user's expectations or requirements concerning data usage for acquiring information in a given area of interest, learning and decision making. Databases should institute quality check measures to ensure that the data they provide to the research community is of high quality. It is much easier to enforce quality measures in a closed setup, but it is a major problem to be addressed in a social collaborative environment (e.g., Wikipedia, 2010[1]). Poor quality data may contain incomplete or missing fields. The data may be represented in a non-standard/legacy formats that will create problems for data and information integration. To ensure high quality, after data is received, databases should then use their own quality measures which may also include manual curation by domain experts. There should be standardized mechanism to ensure consistency and completeness of the submission. Usage of Semantic aware forms (ontology guided forms discussed later) for data procurement, on-the-spot data entry field validations using advance Web scripts may minimize proliferation of inaccurate data.

Data access: Systems researcher often works with diverse set of data usually from different biological levels of organization (molecular, cellular, organism etc). Computational frameworks that will serve to store data and allow data access by query is needed. These computational data framework could give accesses to data by accumulating into one central repository or just through a uniform interface which gives accesses to multiple heterogeneous databases, geographically separated, hosting their own data. With presence of multiple heterogeneous data sources, querying and extracting data will be a problem. Ideal expectation will be a single query to fetch the information spanning several sources. Taking one step further linking biological entities to each other in a meaningfully related manner, enhancing interoperability can be realized by embedding Semantic awareness into the framework. This could enhance the query capacities. Past the query, researchers can retrieve data, compare and analyze until the desired endpoint is attained. This step could be facilitated if analysis and visualization tools are built into the integrated computational framework, allowing users to specify and carry out *in silico* experiments, record intermediate and final results and annotate experiments.

Data visualization: Visualization of raw and modeled data is an important tool for analyzing and interpreting the complex and interconnected data. Visualizing data as pathway and networks has helped researchers to record and communicate their findings. The irony is, in systems research, visualizing deluge of data user may be overwhelmed by it, rather than reaping any benefit at all.

The matter turns to worse when visualization needs to support user interactivity (e.g., in case of assembly, curation and modeling of complex models), and particularly when done in a collaborative setup. So the Web applications and tools supporting visualization should use niche techniques to present the data at the right level of detail, in a cohesive, insightful manner (Gehlenborg et al., 2010). Various way of visually representing the same knowledge breaches effective communication between different biological communities. It takes more effort on biologists to familiarize themselves with different notations. Only recently graphical representation standard for biology, SBGN (Le Novère et al., 2009), has been proposed. However only miniscule of tools exist that has incorporated SBGN (e.g., Cell Designer, Kitano et al., 2005; and SBGN-ED, Czauderna et al., 2010). SBGN awaits adoption by biological tools, databases and Web application. One major problem for SBGN adoption is none of the Web browsers support rendering graphics written in SBGN. Alternatively, Implementing SBGN graphic notations specification in Scalable Vector Graphics (SVG) will be an alternate innovative solution for databases and Web applications (e.g., as proposed by us in CIDMS-PD: Cardiac Integrated Database Management System-Pathway Database, 2009). SVG is a generic graphical representational format has been already widely adopted by internet community (Scalable Vector Graphics, 2010[2]). It is supported and rendered by most of the Web browsers. Community has to yet pick up this idea, and act to develop practical visualization applications.

Data representation and standards: Collaborative nature of systems research place emphasis on conforming to standards and data formats; for searching, information exchange and mutual understanding. Standards can be developed informally among group of researchers or it could be enforced by journals and funding organizations or even by

[1]http://www.wikipedia.org/

[2]http://www.w3.org/Graphics/SVG/

used software tool. It will help to link data and tools into an integrative framework. Standards can also help to avoid misunderstanding and duplication of work. But this will only take off if the community at large can reach consensus on using handful of them. Using agreed set of standards and data formats increases processing efficiency in a large-scale integrative computational framework environment as it minimizes unnecessary, inefficient conversions between standards. Data in the databases and on the Web is a mix of structured and unstructured formats. The representation mechanism is usually simple and diverse. So accessing by machines becomes a fundamental problem.

Security: Communication of data between application systems must ensure security to avoid improper access. *Trust* or the lack thereof, is the most essential factor blocking the adoption of rapidly evolving Web technology paradigm such as software as service (SaaS explained later) and data distribution services. This issue is usually addressed by the database management systems (DBMS) or framework which has mechanisms to handle several security attributes like multi-tenant (clients sharing vital data with servers), data access check and levels of security clearance based on the roles (e.g., admin, general user, curator), data sharing with other organization or participant and keeping the vital data safe from prying eyes.

Version control (VC): Set of mechanisms that support evolution of developed artifacts (e.g., source, analysis and design documents, data, and models) in computer application. VC helps to trackback and in data provenance. Explicitly exposing version metadata to the clients can aid in reinforcing data quality. This issue is usually addressed by the DBMS.

Interoperability: The interoperability issues are the problems that are associated with bringing together heterogeneous and distributed information systems. Today research on interoperability solutions has moved technology from having a single monolithic expensive solution into distributed collaborative inexpensive solutions. With such a trend; often arise the problems of Semantic heterogeneity, data integrity, data representation and data migration, and correctness of the interpretation of data sets obtained from different resources. Semantic heterogeneity, for example deals with the conflict due to multiple names for the same concept used in different resources or it could be as a result of multiple interpretations for the same name. Machines are expected to be told explicitly about such disparities which seem to be very intuitive to humans. Interoperability issues can be attributed to systems heterogeneities occurring at different levels including between softwares, interoperability between analytical methodologies, among data and databases. These problems can be a serious predicament during data integration, analysis and discovering knowledge.

Computationally intensive: Modeling the living organism is a complex task this fact is a reflection of the inherent complexity of biological systems itself. Modeling workflow typically include: defining the problem scope and drawing boundaries in accordance to what questions are needed to be answered; integrating large and diverse data; using integrated data and formalize the problem as model(s) with a machine-readable language; executing the model on the computational infrastructure (including software and hardware); validate, analyze and visualize the results. Varying amount of computing infrastructure is required in every step of the modeling workflow and this requirement increases with more and more complex model (egg. incorporating finer spatial and temporal resolutions into model increases the complexity) (Burrage et al., 2006). All aspects of biological system executes in parallel however computing is sequential. Emulating the parallel processing ability of biological system is a difficult task and not feasible with conventional hardware architectures and existing softwares (Mazza, 2010). They require architecture based on scalable parallel and distributed systems (e.g., grid and Cloud computing, discussed later). Further, to exploit capacities and capabilities of parallel architecture requires advanced software designing methodologies; make difficult paradigm shifts in programming techniques; and implementation of sophisticated algorithm.

Issues with development and distribution of tools: To do meaningful analysis with all the data from various resources requires appropriate tools and methodologies. Typically tools are written with a specific set of requirements and contexts. Systems approach in biology is a rapidly developing field where the pace of data production and progress in methodology is rapid and often there is requirement for new resources. To deal with this situation often there are not many tools available. New tools and standards have to be made or modify the existing ones.

Using domain standards supports interoperability and reuse. Standardization strictly focuses on the most essential and commonalities, but compromises on the variations. In systems driven research if standards are enforced stringently there is a risk that novel findings may be missed. Thus the characteristics of the systems driven research field require adoption of best engineering practices which facilitates development of customized computational infrastructure. The challenge of software customization has been partly met by using off-the-shelf software components that were developed by scientific community during ~omic/post-omic period by several organizations. These readily available software components provide much of the functionalities and capabilities required, which inturn can be chained together in a workflow to achieve much bigger objectives, rather than reinventing the wheel. However, software development using off-the-shelf tools pose several of its own problems than building a system from scratch or scaling a system by re-using components built internally in an organization. The reason is off-the-shelf tools in the first place were meant to run as a standalone application; they have no mechanism for interacting with other programs. It is extremely important, now than ever, the need for collective efforts in the community toward development of infrastructure by creating open-source reusable libraries and toolkits. Coverage of such initiative should not just limit to developing software components but should also be extended for the data, algorithms and analysis methodologies. Although most of these are known issues, the community initiatives to rectify them are progressing slowly because of political, funding and intellectual property reasons.

APPROACHES, TECHNOLOGIES, ARCHITECTURE AND DESIGN STRATEGIES FOR DATA INTEGRATION

Systems research is highly interdisciplinary and involves meaningful interpretation of data from high-throughput experiments through building multiscale models. There is a continuous need to integrate existing technologies with newly developed and emerging technologies. In this section we discuss systems research driven design and developmental strategies undertaken toward data integration by building databases and Web application infrastructures. Further we will discuss how such an information infrastructure can allow disparate research groups to access integrated data sources, reuse tools and methodologies that help cross-collaboration in generating data and models. Our main focus here would be to summarize database and Web application technologies.

The approaches to integrate data can use centralized model or distributed model (Sheth and Larson, 1990). In centralized model there is one unified schema, a massive central repository (e.g., warehouse), which is framed based on the schemas of the individual

data sources (Reddy et al., 1994). The data transferred to central repository is collected, integrated, stored and made available for search and presentation. (e.g., Biowarehouse, Lee et al., 2006; Atlas, Shah et al., 2005). The distributed model includes federation and mediation approaches. In federation approach the data is left in the respective fully functional expert databases maintaining data autonomy while still providing integrated access to distributed data (e.g., Entrez, Sayers et al., 2010; Biomart, Haider et al., 2009; DAS, Jenkinson et al., 2008; EBI, Brooksbank et al., 2005). Here integration expects no data transfer to any one central repository. The design relies on an agreed data exchange protocol between the participating databases. A central hub undertakes responsibility of coordinating and organizing the queries across databases and data retrieval is powered by each databases. Mediation does not store any data on its own rather it provides virtual view of the integrated sources (e.g., DiscoveryLink, Haas et al., 2001).

All the integration solutions till date can be grouped in to three different technological layers: Data layer centric solutions, middle/object layer centric solutions and application layer centric solutions. The data layer centric solutions involve databases and DBMS in the form of data warehouse, multi-databases, distributed databases or federated databases. The middle/object layer centric solutions mainly support distributed applications. Many technologies that belongs to object/middle layer centric solutions are related to middle ware development. Middleware is typically software that resides between a data store on one side and applications on the other where data is collected or processed further. The object based approach is fixated on use of interoperable standard objects. Examples of technologies belonging to this category are CORBA (Common Object Request Broker Architecture), SOAP (Simple Object Access protocol), SOA Service-oriented Architecture, DCOM (Distributed Component Object Model), Representational State Transfer (REST) and Java (EJB, RMI). In the application layer centric solutions applications takes responsibility to integrate data. Link integration, view integration and Web services (a variant of link integration refer to review by Stein, 2003) are approaches that belong to this layer. The projects utilizing application layer centric solution have used centralized model, distributed model and some have even used a hybrid of the two models (e.g., CIDMS, 2007; ApiDB, Wang et al., 2007), which eliminates several disadvantages (e.g., helps to improve performance) posed by either of the models.

CENTRALIZED DATABASES

One of the early popular data integration technique using centralized model was by providing unified interface to heterogeneous data sources (e.g., SRS, Sequence retrieval system; Zdobnov et al., 2002), List of publicly accessible SRS servers (Biowisdom, 2010). Central repositories are created locally by full-text indexing the data present mostly in the form of flat-file/XML format (locally mirrored data sources). Some of these repositories even allow seamless integration with numerous bioinformatics analysis tools. The users can use keywords, identifiers like accession and symbols to search and navigate through data contained in various databases regardless of their format; query them in same way, at the same time and capture results. One of the variation of the above approach is mining descriptors (representative data) from various databases based on some predefined criteria (e.g., Gene Cards database organizes one file per human gene (Safran et al., 2002), organize them as files which then will be collectively indexed and enabled to full-text search. The descriptors obtained from the sources will contain only the most essential information and hyperlinks to original source. User will be presented a collective summary from multiple sources in a single information space with search and data filtering capabilities. The drawbacks are that it is incapable of supporting searches based on Semantics that uses hierarchically structured information (Ontologies).

One of the widely used technologies based on centralized model is Data Warehouse (e.g., Biowarehouse, Lee et al., 2006; GUS, Clark et al., 2005 and Atlas Shah et al., 2005). In this approach data is imported from all remote sources via special scripts/programs called loaders into one single local database. The loader is a piece of software that helps in conversion of data in a different format to the required format. The loaders can also be designed to apply a degree of semantic normalization to their respective source data, decreasing semantic heterogeneity (discussed later). The information imported from various databases is collected and organized in a unified data model (Lee et al., 2006). This way of data integration provides a single access point to a collection of all data with capabilities of answering the questions; not just those of which individual source database could have answered but also answers to other complex questions that requires integrated information which none of individual source database could have provided. Other key benefits include good performance and improved data consistency. Among the major problems with this approach include keeping information up to date (data synchronization), scalability (which involves tinkering with database schema and writing/changing existing loader programs) and data privacy. Despite these disadvantages several recent projects (RDFScape, Bio2RDF, CardioSHARE, KNO.E.SIS) in life science domain which is based on the state of the art Web technology, the Semantic Web technology, (discussed later) uses centralized model mainly because of the performance limitations in federated approach.

FEDERATED DATABASES AND WEB SERVICES

The problems of data warehouse approach are resolved in federated databases, mainly because centralization of data is not a necessity. Federated databases are playing increasingly large role in life science data integration and several databases/Web applications projects have embraced this approach.

With adoption of distributed model it is implicit to expect scattered heterogeneous data resources. There is a need for a technology that is able to automate access to remote resources, manage and link data properly. Web service is such a technology which is employed to address the issues of distributed model mainly concerning application to application communication. A programmatic interface to a resource facilitating application to application communication made available over Web is often referred to as Web service (W3C, 2002).

Web services technology uses SOA. SOA architectural model decouples service provider (source) from the service consumer (sink). The goal is to provide a great flexibility in constructing distributed computing systems based on services. This means that service consumer can choose any service from any provider no matter which language is used for its implementation and what platform

they run on, as long as the interface is compatible. XML is accepted as a ubiquitous representational language for data integration and interoperability (Achard et al., 2001). For the same reason XML-based standards (e.g., SOAP and Web Services Description Language, WSDL) are predominantly used for describing data, services and the communication protocol maintaining interoperability between services (Neerincx and Leunissen, 2005). The implicit advantage is that the decoupled nature of the approach provides a means to develop solutions that could keep pace with rapid and dynamic developments in systems biological research. In this setup softwares can evolve separately, made interoperable, easily implemented and scaled (e.g., DAS, Jenkinson et al., 2008; Hmida et al., 2005). Because of these advantages in terms of flexibility and extensibility many of the biomedical databases have started providing Web service (e.g., NCBI, Sayers et al., 2010; EMBL, DDBJ, BioMoby, Wilkinson and Links, 2002; caGRID, Saltz et al., 2006, pathway commons, Cerami et al., 2006; Biomodels, Li et al., 2009).

Web services have paved in the evolution of tools which could help to: (1) display and access to integrated content on Web/application interfaces (e.g., jemboss, Carver and Bleasby, 2003; SeWeR, Basu, 2001; cPath, Cerami et al., 2006; e.g., link and view integration, Stein, 2003); (2) render complex biological interactive visualizations (pathwayExplorer, Mlecnik et al., 2005); (3) automation of Interactive forms to accept data from user (e.g., Xforms, W3C/CWI, 2010); and (4) most importantly development of dynamic network of XML-based data pipelines which could be used by analytical tools (e.g., CellDesigner plugins, Funahashi et al., 2007; Van Hemert and Dickerson, 2010; Cytoscape plugins, cPath[3]; WikiPathways, 2008), including development of advanced suits for automatic workflow generation (e.g., BioMOBY, Wilkinson and Links, 2002).

In a distributed model, researchers are often needed to access several services to accomplish a useful task. Often researchers face interoperability of the services as a major problem. In response, they resort in creation of their own workflow by fetching data from one source, usually reformat it, submit it to a service of another source, parse the results, reformat again and resubmit. This endeavor will continue till an acceptable end result is accomplished. Many projects have tried to solve the interoperability problems by developing specialized platforms called grids (e.g., caGRID, Saltz et al., 2006; PathGrid, Arbona et al., 2007; Walton et al., 2010; The Virtual Kidney, Harris et al., 2009; Abramson et al., 2010; GEMSS, Benkner et al., 2005). At its conception the "Grid" was envisioned as a distributed and cost-effective solution to boost computational power to solve large-scale mathematical and data-bound problems. Current mature understanding of the grid is more as a robust framework (mostly based on principle of SOA) for performing distributed computing tasks on the scale of internet which can enable service-oriented science (Foster, 2005). The gird services can provide either a HTML based interface (classic grids) or much advanced SOAP interface. SOAP interface is widely accepted, because of several advantages over classic grids (Neerincx and Leunissen, 2005). In a grid, services are distributed over many servers, and clients use specialized software to discover and execute these services. Usually a grid uses middleware that uses wrappers around existing programs to create a standard application programming interface (API) for communication between services. Number of useful user-friendly tools have been developed to support the grid platform including graphical workflow management tools (e.g., Taverna, Oinn et al., 2006), schedulers (e.g., GridFlow, Bo et al., 2005), and script translation tools (GridAnt, Amin et al., 2004; SQUID, Carvalho et al., 2005).

Another technology which is used for building distributed systems is CORBA. CORBA is tightly coupled, object centric and stateful. In comparison, Web services are loosely coupled, utilize a message exchange model and are stateless. This difference gives Web services a flexibility and simplicity to implement distributed system which is not seen in CORBA implementation. However CORBA may interoperate and coexist with Web services similar to grid. That means CORBA has a similar architecture as Grid to work with Web services, but important distinction is that Web services are integral part of grid, while it is not a native component of CORBA.

Another important advantage of grids is its ability to leverage on existing IT infrastructure to optimize usage and sharing of computational resources and manage large amounts of data. The rationale behind grid technology is similar to eclectic power grid where users do not know the details of the technology and the sources. They simply connect to common interfaces, subscribe and consume what they need.

One of the main problems with grid computing infrastructure is over-provisioning of computational resource. For this reason grids are suited for data-intensive task and would not be economically feasible for small tasks. Recently Cloud computing has evolved from grid computing and provides on-demand resource provisioning (Protein Bioinformatics Infrastructure for the Integration and Analysis of Multiple High-Throughput "omics" Data[4]). Readers are encouraged to refer to reviews by Bateman et al. (Bateman and Wood, 2009; Martin, 2001) and Stein (2010) for detailed discussion of cloud computing and its potential applications in biological field. The consumers of cloud service need not own the infrastructure, software or platform in the cloud and need not care about how they are maintained. The scaling of computational resources is dynamic (on demand) and easy because of virtualization technology. For these reasons like many others, we expect that a wide adoption of cloud computing for systems research wanting varying demand for computational resources, and cost effectiveness. Already we can witness its adoption in few of the genomics initiatives (Langmead et al., 2009; Baker, 2010; Schatz et al., 2010).

Availability of SOAP/WSDL based Web services merely gives programmatic access to databases and Web applications. SOAP and WSDL alone is incapable of self-describing the services and data, thus machines can not anticipate the meaning of appropriate services and their interfaces. This is a major problem for data integration using Web services because of Semantic heterogeneity. Semantic heterogeneity is caused by disagreement about the meaning, interpretation, or intended use of the same or related data and services. Since Web service architecture lacks semantics realizing functionalities like automated service discovery, mediation and reuse of services is not possible. These functionalities are essential for linking Web services and creating service pipelines, for enabling efficient and more meaningful use of Web

[3]Cytoscape cPath PlugIn

[4]http://www.hindawi.com/journals/abi/2010/

services. Such pipelines help users to explore and manipulate data, automate analysis and solve complex biological tasks. In response ontology based solutions (Semantic Web services) were developed and are used in several successful projects [e.g., caCORE, Komatsoulis et al., 2008; BioMOBY(S-MOBY), Wilkinson et al., 2005; myGRID workflows using BioMOBY services, Kawas et al., 2006; DiscoverNet, Ghanem et al., 2002 and TAMBIS, Stevens et al., 1999]. Projects that couple semantics with the Web services functions more effectively as it eliminates critical interoperability problems, which commonly surfaces when only Web services are used. They define a registry-based discovery system based on Universal Description, Discovery and Integration protocol (UDDI, is a standard by OASIS standard consortium, to create service directories that enable applications to dynamically find and use Web services UDDI[5]). Each of the projects listed here addresses the interoperability problems using Semantic Web services differently. We encourage you to refer to a review by Good and Wilkinson (2006) which describes the technology and motivation behind using Semantic Web services in some of the above projects.

Performance of Web service is also a concern confronted especially integrating large dataset. The performance overhead is introduced by SOAP message size (XML text) and complexity, XML parser, cost of serialization and deserialization, cost of connection establishment, security validation, UDDI registration and querying of XML. One of the efficient ways to improve performance of Web services is by minimizing communication delays. This could be achieved by using compression formats like binary XML which has shown to provide performance boost. Another societal solution to this end is by community agreeing on one data representation format which will prevent unnecessary, inefficient conversions between formats. For example pathway data available from public databases made available primarily in one of the following formats: cellML, SBML, and BioPax. To integrate all the available pathways one has to resort to inefficient conversions (using converters like CellML2SBML, SBML2BioPax) between these formats. This especially in real time is overkill.

ONTOLOGY AND SEMANTIC WEB TECHNOLOGY (SWT)

Semantic Web Technology deals with the meaning of information, enables computers to understand the Web content, perform tedious task like finding and assembling knowledge from multiple sources on the Web. In the above sections we have already discussed the use of SWT in context of discovery of Web services. The most important scenario where Semantic Web matters is for identifying semantically related (having same meaning) concepts from different resources. Having a mapping between related concepts in database will help to query multiple databases with a single query. Also it should be possible to automatically identify and map the various data fragments creating rich information spaces that can be explored for new knowledge.

The SWT thus focuses on using: (1) ontologies to explicitly specify the domain concepts; (2) standard representation languages [e.g., Resource Description Framework (RDF), 2010[6]; RDF schema (RDFS), RDF Vocabulary Description Language 1.0: RDF Schema (RDFS), 2004[7]; Web Ontology Language (OWL), 2004[8]] to name, encode, describe, combine information; (3) standard Web protocols to access the information (e.g., Query language for RDF, SPARQL, 2008[9]); and (4) technologies to leverage on computational task like inference and distributed query (inference engine, Ruttenberg et al., 2009). We list here several interesting reviews (Tyrelle, 2005; Wang et al., 2005; Good and Wilkinson, 2006; Post et al., 2007; Ruttenberg et al., 2007; Sagotsky et al., 2008; Antezana et al., 2009) covering in detail about SWT.

Ontology is defined as a shared vocabulary plus a specification of its intended meaning (Guarino, 1998). Thus ontologies unambiguously represent concepts that are known and are necessary in Semantic Web for resolution of naming conflicts. Since SWT is depended on the ontologies there is a need to rapidly develop globally accepted quality ontologies. Several ontologies [e.g., Gene ontology (GO), Ashburner et al., 2000; Cell cycle ontology, mammalian Phenotype ontology, Antezana et al., 2009; SBO, Le Novere, 2006; BioPAX, 2006; pathway ontology 2010[10]; Event ontology, Kushida et al., 2006] now are integral part of any biological and systems driven research. The details of each of these ontologies and many others can be found on Global open biological ontologies Web site (The Open Biological and Biomedical Ontologies[11]). Alternatively use Bioportal (Noy et al., 2009), a Web repository for biomedical ontologies to access and share ontologies. We request you to refer to useful reviews (Bard and Rhee, 2004; Jurisica et al., 2004; Puustjarvi and Puustjarvi, 2009) on ontologies for knowledge management in biomedical field.

The Semantic Web has adopted basically three formal languages that are based on the mathematical graph model and are machine readable: (1) RDF usage makes information essentially self-describing. RDF represents data by making statements called "triples" in the form of a subject-predicate-object. Triples are analogous to a "complete sentence" consisting of subject and predicate (fine verb and object), which forms basic building blocks of expression in natural language. Collection of sentences can make paragraph, similarly, collection of triples (even spanning multiple documents) can form networks of interconnected logical graphs that describe information nodes and their interrelationships with other nodes essentially integrating information. Also OWL is built on RDF; (2) RDFS, is a framework overlaid on RDF graph model to specify a standard way to describe a resources represented in RDF, with in a particular domain. While RDF provides model and syntax for describing resources, but by itself is incapable of defining the meaning of those resources. For example RDFS can be used to create vocabularies to describe anything from diseases to molecules, experiments, instrument to even abstract concepts like consciousness and behind; and (3) OWL is used to define content of the information by defining the types of objects, their vocabulary and their relationships in an RDF document. OWL facilitates greater machine interpretability of Web content than that supported by

[5]http://www.uddi.org423589.cta.html

[6]http://www.w3.org/RDF/FAQ

[7]http://www.w3.org/TR/rdf-schema/

[8]http://www.w3.org/TR/owl-features/

[9]http://www.w3.org/TR/rdf-sparql-query/

[10]http://sourceforge.net/projects/pathwayontology/

[11]http://www.obofoundry.org/

XML, RDF and RDFS by providing extended vocabulary along with formal semantics. OWL has three increasingly expressive sublanguages: OWL Lite, OWL-DL (description logic), and OWL Full. Basis for OWL-DL is first order description logic. This means that OWL-DL ontology is expressed in formalism with well-defined semantics and over which automated reasoning can be undertaken. In our view, OWL is ideal language to capture knowledge in terms of Ontologies (Stevens et al., 2007) and can satisfy most of the requirements in systems driven research.

Semantic Web languages rely on unique identity (a global identifier) through use of Uniform Resource Identifiers (URI, URI Interest Group, 2005[12]). URI is a string of characters used to identify a name or a resource on the internet. So URI identifies a resource either by location (Universal Resource Location, URL) or name (Universal Resource Name, URN). Such identification enables interaction with representation of the resource over a network (World Wide Web) using specific protocols. Also, it is by mapping URI for an object in two separate Semantic documents (RDF documents) one is able to integrate the information together. Thus SWT using URIs can help us to overcome data integration problems with fewer efforts.

Semantic Web Technology are increasingly gaining acceptance in biological community and several projects have spanned like Bio2RDF, RDFScape (Splendiani, 2008), YeastHub (Cheung et al., 2005), BioPAX and semanticSBML (Krause et al., 2010) to name the few. More and more biological data and ontologies are now made available in RDF (e.g., AlzPharm, Lam et al., 2007), UniportRDF and OWL (e.g., GOOWL, Ashburner et al., 2000; biOzen, UMLS ontology MGED OWL). To support progress of Semantic Web technology, several softwares are developed by open-source developers and made available through a centralized Web portal SemWebCentral[13]. Some of the open-source efforts that we would particularly point you to is for storage and retrieval solutions for RDF triples supported by SPARQL for querying (e.g., Virtuoso, Sesame, 3store, Harris and Gibbins, 2003; Mulgara) and frameworks for building Semantic Web applications including rule-based inference engine (e.g., Jena, MobyServlet, Gordon et al., 2007).

TECHNOLOGIES WITH SOCIAL NETWORKING: WEB2.0

Technologies of data integration discussed till now uses data exchange between computer and its user. Concept of data sharing and collaboration among user is not part of such integration facility. Web2.0 is the technology based on user-computers-user collaborative model, and transcends traditional data integration technologies; giving it an edge to become a more suitable platform for enhancing systems research. The Web2.0 is conceived as a social, collaborative and collective Web space (Kamel Boulos and Wheeler, 2007; Zhang et al., 2009). Unlike Web1.0 which is "read only;" merely meant to display information on the Web, Web2.0 and 3.0 is "read and write," where user constantly interacts with the Web and works in a networked setting.

Key characterizing elements of Web2.0 are social Web, user added value, use of Web service and software as service (Zhang et al., 2009). The Social Web and user added value emphasizes on the need to connect people and use the collective power of community to achieve data integration tasks. Wikipedia is an ideal example for a successful Social Web project in which the content is both created and edited by users. Owing to wikis success there are considerable efforts in biological community for porting Wikipedia into biological domain (e.g., Gene Wiki, Huss et al., 2008; WikiProtein, Mons et al., 2008; Wiki Pathway, Pico et al., 2008).

In order to facilitate interaction between the user and the computers on the Web, both data and interface used by user should be dynamic. Two of the most powerful programming techniques which have facilitated such interactivity on Web (by building rich client applications) are AJAX (Asynchronous JavaScript and XML) and Flash. Both AJAX and Flash are equally good for creating interactive Web application. We and similarly many biological projects have used AJAX in several of our projects basically because it is an open-source initiative and it by default is supported by almost all the Web browsers. In a typical Web technology client makes a request to server and waits for the return of response by the server and then takes some action on the content sent, before client can make another request. Unlike, using AJAX clients are allowed to continuously interact with the server without having to wait for the immediate response to return for each request made. The processing of transactions happens in background which facilitates exchange of messages between client and server without any interruptions. Web application designed on such rich interactivity are increasingly used for presentation of data, rendering complex and dynamic visualizations, context dependent user-friendly search and browsing interfaces and developing context data submission and feedback forms. One of the recent project that could be valuable for systems biology community is Payao (Matsuoka et al., 2010) built using Flash. Payao is a community-based, collaborative Web service platform for model curation, tracking updates and tagging system. Users can collaboratively engage in model building and curation processes. Payao supports standard representations like SBML (Systems Biology Markup Language, Finney and Hucka, 2003) as input and output data format and SBGN (Systems Biology Graphical Notation, Novère et al., 2009) for visualization of the model.

Technological implications and advantages of Web services as a means to support computer-to-computer interaction were discussed earlier. Web2.0 caters to extensive use of Web services to develop interoperability between data resources and software through exposing Web application programming interfaces (Web APIs) described using WSDL. Providing Web services based pipeline will help users to explore and manipulate data, mix and match Web services that can use the data in variety of novel ways and solve complex problems. Launching several useful tools and algorithms that are often used for analysis, data management and visualization similar to Google docs, 2010[14], as Web applications and services (software as a service, SaaS) will help the scientific community in multiple ways. It is better and advantageous way, doing work online constantly connected, collaborating and sharing information. Software can be constantly improved in response to user

[12]http://www.w3.org/2001/12/URI/

[13]http://www.semwebcentral.org/?page_id = 12

[14]https://www.google.com/accounts/ServiceLogin?service=writely&passive=1209600&continue=http://docs.google.com/&followup=http://docs.google.com/&ltmpl=homepage

feedback and needs. It also eliminates software platform dependency enabling use on diverse operating platforms (e.g., UNIX, Mac, Windows and Android) and need tedious task of local installations. All this eases data assembling and integration from heterogeneous data sources and as a result Web2.0 is likely to promote discovery of new knowledge.

FUTURE OPPORTUNITIES, NEW CHALLENGES AND RECOMMENDATIONS

The role of modeling and simulation in the Systems driven analysis of living systems is now clearly established. Emerging disciplines, such as systems biology, and other worldwide research initiatives, such as the Physiome project (Hunter et al., 2008) and the Virtual Physiological Human project (Fenner et al., 2008; Hunter et al., 2010), are based on an intensive use of modeling and simulation methodologies and tools. One of the key aspects in this context is to perform an efficient integration of various models representing different biological/physiological functions, at different levels of organization spanning through different scales. To handle such complex integration challenges and improve our ability to conduct biologically meaningful system scale analysis require a unique, interoperable, universal information framework with the following characteristics (Boyle et al., 2008): (1) Unique identification and dynamic data resolution with capabilities to track data provenance; (2) Services to manipulate data (e.g., relationship services, synonym services, query service, registry service, ontology service); (3) Services to analyze the data (e.g., to run inferential analysis, statistical and mathematical analysis, simulation services); (4) Services for data presentation and visualization (e.g., rendering complex interactive networks and pathways, online collaborative modeling); (5) Semantic Web enabled (common data syntax, shared semantics and Semantic discovery) with capabilities to access data and Semantics using same mechanism (e.g., dynamic discovery of services); and (6) implementation should be sufficiently robust and portable to allow use by researchers with a wide variety of backgrounds and computing expertise. A data/Web infrastructure using such frameworks will be scalable with efficient data handling potential. Number of novel solutions for developing information frameworks with the above functionalities has been proposed in terms of prototype systems (e.g., Simple Sloppy Semantic Database (S3DB); Almeida et al., 2006), design methodologies for databases (Maier et al., 2009), and software (e.g., I-cubed, Boyle et al., 2008; generative software development, Nord and Czarnecki, 2004). However as of now, no single framework is implemented that can support all/most of the listed functionalities.

Like many other research groups, we also agree that future of systems research requires semantically based data integration through ontologies. Putting data into easily accessible repository in the standardized format is essential part of realizing Semantic Web vision. There is a need for creation of tools, databases and Web applications that makes creating, publishing and searching of RDF/OWL intuitive and simple for biomedical researchers and clinicians. To this end existing databases should actively participate in representing their data in Semantic languages like RDF/OWL and provide Web services. The potential benefits of SWT are being realized increasingly in life science and health care community. There is work in progress trying to address the issues, develop, and support the use of SWT through internationally organized efforts (e.g., Semantic Web Health Care and Life Sciences Interest Group, HCLSIG[15] initiated by W3C).

Just as any evolving new technology, Semantic Web is also full of issues. Two particular prevailing problem of Semantic Web is the ambiguous identification of resources and vagueness of resource definition (Wang et al., 2005). This means that different ontologies often refer to the same concept with different URIs and a particular resource could have ambiguous descriptions. This issue of ambiguity calls for URI resolution steps, while integrating data. Since biologists who are involved in ontology development are focused on Semantic rather than the Web technology. Just creation of ontologies will be of little use if developed ontological concepts are kept hidden (inaccessible) or redundantly identified in the Semantic Web. It is extremely important to provide explicit access to each ontological concept via resolvable URIs. Either ontology providers should build concept resolution system by themselves or they should make the ontologies available to consortiums that have infrastructure to launch ontology resolution services. Also URI harmonization strongly requires both technical and social collaboration. Currently several proposal for standardized identification of biological entities and relationship (e.g., life science identifier, LSID; Martin et al., 2005; URI based, URI Interest Group, 2005[12]; MIRIAM URIs, Laibe and Le Novere, 2007) are put up. However there is no consensus yet reached to use a common accepted identification scheme. The Semantic community should work toward a possible agreement of using an explicit identification system that could help unambiguously specify biological resources. To address this problem, efforts by Shared Names initiative (2009)[16] is a valuable beginning to normalize URI's in the biomedical context. Having met the above requirements, in future, SWT will receive wide adoption with in systems driven research community.

Current representative applications of Semantic Web are SWEDI (Post et al., 2007) and an example system developed by HCLSIG[15]; at present they both manage issues of data integration but not much work is undertaken implementing data analysis and interpretation functionalities. The reason lies in practical difficulties implementing such mechanisms to extract meaningful knowledge from raw integrated data. Very few recent projects have tried to leverage the capabilities of inference technology. The RDFScape (Splendiani, 2008), is one of the few recent work in life sciences domain, which has attempted to use reasoners (program which can determine relations among ontology classes) on BioPAX (2006) data inside Cytoscape (Shannon et al., 2003). In another novel study a different approach using Semantic Web methodologies to integrate gene data with phenotype data was demonstrated. It used RDF graph (network) analysis with reasoners to prioritize candidate cardiovascular disease genes (Gudivada, 2007). However all the projects that have used inference technology are tried on the data integrated from limited number of resources, but not on Web-scale datasets.

Ideally where RDFS and OWL constructions are used, it should be possible to apply automated reasoning over data schema and innovate meaningful knowledge. However at present Semantic Web is not completely ready to equip with inference engines. One of the potential reason is problems posed by large ontologies [e.g., Unified Medical

[15]http://www.w3.org/2001/sw/hcls/

[16]Available on: http://sharedname.org/page/Main_Page

Language Systems (UMLS), 2004[17], NCI thesaurus (NCIt, 2010)[18], GO, Ashburner et al., 2000; Good and Wilkinson, 2006]. It is currently unfeasible to retrieve, modify and process concepts at runtime as conceived for their utilization on Semantic Web. Because current reasoners and other tools that support Semantic Web require that all the information that they process should be loaded into memory. This can severely curtail performance or even fail when scaled to large ontologies. Several algorithms have been proposed to decompose large ontologies into less manageable and meaningful pieces retaining some of the semantics of the full version. Researchers can then further modify these ontology models to their specific needs (for example reducing the complexity and/or cleaning up inconsistencies) creating an inference feasible versions. However multiple limitations still persist, requiring future research aimed in this direction.

In previous section (Federated Databases and Web Services) we discussed how Web services can help to develop federated systems that could keep pace with the rapid advances in systems research. But, numerous issues associated with Web services could hinder the progress. Issues related to maintenance of code that might affect the scalability and ease of development when Web services are built on SOA. Newer software development paradigm like Aspect Oriented Programming (AOP) solves the problems associated with code-tangling and code-scattering (Kiczales et al., 1997). But, adoption of AOP is yet to percolate into life science development stream.

Ontologically described Web service interfaces are not yet completely available which need to be addressed for realization of automatic discovery of services. A limitation of WSDL and SOAP is it being purely syntactical cannot express the semantics of underlying data and services which renders them inaccessible by machine. Adding semantics to represent the requirements and capabilities of Web services is essential for achieving unambiguity and machine interpretability. Work on automatic higher level integration of Web services and data by machine is in its incipient stage and progressing slow. The reason underlying slow implementation of such a useful infrastructure is: (1) it presupposes a presence of formal logic over Web resources (i.e., Web represented in semantic languages, McIlraith et al., 2001); (2) at present Web services uses WSDL, which lack a way to describe semantics. Recently, WSDL-S (semantic markup of Web services description language) was proposed as an alternative solution to the problem (Miller et al., 2004); and (3) it could be attributed to lack of efficient Web crawlers to index Web service description analogous to what Google does for Web pages. The value additions that a Web service search engine can provide could be witnessed in BioCatalogue (Goble and De Roure, 2008) and BioMoby (Wilkinson et al., 2005) projects that are manually curated search engines for Web services.

Adoption of Web3.0 will revolutionize the way we manage data online, exchange information with each other and discover knowledge from rapidly accumulating biological data through collective intelligence. The Web3.0 is defined by Wiktionary as: "The predicted third generation of the World Wide Web usually conjectured to include semantic tagging of content" (Wiktionary: a wiki-based open content dictionary, 2010[19]). This means Web3.0 is an extension of second generation Web2.0 technology and is semantic aware Web technology with each service being closely coupled with a formalized description. The range of tools (inference searches, ontologies, SPARQL) envisioned as part of Semantic Web will be available in Web3.0. Conceptually in Web3.0 entire Web is viewed as one large integrated database. Allowing structured information to be read by different programs across the Web and enabling users to do more accurate searches and finding precisely what they want.

Scientific workflows are ideal for *in silico* experimentation in advancing systems research. Several of the workflow systems are significantly developed including the Taverna workflow workbench (Oinn et al., 2006), Kepler (Ludäscher et al., 2006), Triana (Taylor et al., 2007), and Pegauss (Deelman et al., 2005). Workflows include number of master services described in WSDL file that coordinate or aggregate activities together. Some of the workflow can be highly scalable to span multiple domains and organizations dispersed in different geographical locations. In such a scenario user interactions with the workflows at several intermittent levels is preferable facilitating interactive steering and monitoring. This will give user, control over exception handling, monitoring data, and choosing alternative work paths (steering) depending on the results witnessed at runtime. Workflow management will be especially crucial for computational intensive and long-running workflows and services which are typically encountered in *in silico* systems scale analysis. Further, Business Processes Enterprise Language (BPEL) was recently proposed; it has several key advantages to specify scientific workflows in a distributed computational setup (Akram et al., 2006; Tan et al., 2010).

To succeed, systems research must be a collaborative, cross-disciplinary and a broad organizational endeavor similar to successful initiative like Alzforum (Lam et al., 2007), MyExperiment (Goble et al., 2010), and Payao (Matsuoka et al., 2010). Several tasks integral to systems research (such as ontology development, URI standardization, developments of tools) requires involvement of scientists from different backgrounds including biologist, physicians, computer scientist, mathematicians and statisticians. To facilitate this type of multidisciplinary interaction, certain prevailing challenges must be met including: (1) adoption of machine-readable data representation formats including semantically aware formats; (2) workflows to address data quality and integrity; (3) implementation of resource identity; and (4) tracking of provenance and ownership. The vision of the Web2.0 and particularly Web3.0 wraps these ideas in to framework of Semantic Web. Useful collaborative Web tools and applications like wiki, blogs, mashups, and light weight Web apps for integration of distributed Web resources on demand will be made available for systems driven research community fostering active participation and an opportunity to take advantage of its integrative and analytical potential.

CONCLUDING REMARKS

We have discussed the current state of the art approaches and technologies and open issues of database and Web application implementation in context of systems driven research. First we provided issues associated with the data integration and then we discussed how these issues have been tackled. And, finally, we discussed the corresponding open issues and their possible solutions. Despite considerable progress in appropriate technologies and efforts to establishing an efficient computational platform, the integration of biological data to meet systems driven research will remain a challenging problem for both present and conceivable future. We need to stay attuned

[17]http://www.nlm.nih.gov/research/umls/about_umls.html

[18]Available on: http://ncit.nci.nih.gov/

[19]http://en.wiktionary.org/wiki/Web_3.0

to three important aspects that drive the field: science, technology and society. Only by a consorted effort and support by all players of research community like database providers, funding agency, experimental and theoretical biologist, we will be able to bring revolution in systems driven research. To this end selecting and implementing the most appropriate technology is of paramount importance.

ACKNOWLEDGMENTS

This work was supported by the following grants: (1) the NRF grant (2010-0002159) funded by Korean Ministry of Education, Science and Technology (MEST) and (2) the Gwangju Institute of Science and Technology (GIST) Systems Biology Infrastructure Establishment Grant (2010).

REFERENCES

Abramson, D., Bernabeu, M. O., Bethwaite, B., Burrage, K., Corrias, A., Enticott, C., Garic, S., Gavaghan, D., Peachey, T., Pitt-Francis, J., Pueyo, E., Rodriguez, B., Sher, A., and Tan, J. (2010). High-throughput cardiac science on the Grid. *Philos. Transact. R. Soc. A Math. Phys. Eng. Sci.* 368, 3907–3923.

Achard, F., Vaysseix, G., and Barillot, E. (2001). XML, bioinformatics and data integration. *Bioinformatics* 17, 115–125.

Akram, A., Meredith, D., and Allan, R. (2006). "Evaluation of BPEL to scientific workflows," in *Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*. IEEE Computer Society.

Almeida, J. S., Chen, C., Gorlitsky, R., Stanislaus, R., Aires-de-Sousa, M., Eleuterio, P., Carrico, J., Maretzek, A., Bohn, A., Chang, A., Zhang, F., Mitra, R., Mills, G. B., Wang, X., and Deus, H. F. (2006). Data integration gets "Sloppy". *Nat. Biotechnol.* 24, 1070–1071.

Amin, K., Von Laszewski, G., Hategan, M., Zaluzec, N. J., Hampton, S., and Rossi, A. (2004). "GridAnt: a client-controllable grid workflow system," in *Proceedings of the Hawaii International Conference on System Sciences*, Big Island, Hawaii.

Antezana, E., Egaña, M., Blondé, W., Illarramendi, A., Bilbao, I., De Baets, B., Stevens, R., Mironov, V., and Kuiper, M. (2009). The cell cycle ontology: an application ontology for the representation and integrated analysis of the cell cycle process. *Genome Biol.* 10, R58.01–R58.19.

Antezana, E., Kuiper, M., and Mironov, V. (2009). Biological knowledge management: the emerging role of the semantic web technologies. *Brief. Bioinform.* 10, 392–407.

Arbona, A., Benkner, S., Engelbrecht, G., Fingberg, J., Hofmann, M., Kumpf, K., Lonsdale, G., and Woehrer, A. (2007). A service-oriented grid infrastructure for biomedical data and compute services. *IEEE Trans. Nanobiosci.* 6, 136–141.

Asai, Y., Suzuki, Y., Kido, Y., Oka, H., Heien, E., Nakanishi, M., Urai, T., Hagihara, K., Kurachi, Y., and Nomura, T. (2008). Specifications of insilicoML 1.0: a multilevel biophysical model description language. *J. Physiol. Sci.* 58, 447–458.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–39.

Baker, M. (2010). Next-generation sequencing: adjusting to data overload. *Nat. Methods* 7, 495–499.

Bard, J. B., and Rhee, S. Y. (2004). Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.* 5, 213–222.

Basu, M. K. (2001). SeWeR: a customizable and integrated dynamic HTML interface to bioinformatics services. *Bioinformatics* 17, 577–578.

Bateman, A., and Wood, M. (2009). Cloud computing. *Bioinformatics* 25, 1475.

Benkner, S., Berti, G., Engelbrecht, G., Fingberg, J., Kohring, G., Middleton, S. E., and Schmidt, R. (2005). GEMSS: grid-infrastructure for medical service provision. *Methods Inf. Med.* 44, 177–181.

BioPAX (2006). *Pathway Exchange Language for Biological Pathway Data*. Available at: www.biopax.org

Biowisdom (2010). *Public SRS Installations*. Available at: http://www.biowisdom.com/download/srs-parser-and-software-downloads/public-srs-installations/

Bo, C., Liu, Q., and Yang, G. (2005). Distributed gridflow model and implementation. *Lect. Notes Comput. Sci.* 3379, 84–87.

Boyle, J., Cavnor, C., Killcoyne, S., and Shmulevich, I. (2008). Systems biology driven software design for the research enterprise. *BMC Bioinformatics* 9, 295. doi: 10.1186/1471-2105-9-295.

Brooksbank, C., Cameron, G., and Thornton, J. (2005). The European bioinformatics institute's data resources: towards systems biology. *Nucleic Acids Res.* 33, D46–D53.

Burrage, K., Hood, L., and Ragan, M. A. (2006). Advanced computing for systems biology. *Brief. Bioinform.* 7, 390–398.

Cardiac Integrated Database Management System-Pathway Database (CIDMS-PD) (2009). *An Innovative Systems Biology Knowledgebase for Cellular Pathways in Heart*. Available at: http://www.icsb-2009.org/schedule_details.php?ID=272

Carvalho, P., Gloria, R., de Miranda, A., and Degrave, W. (2005). Squid – a simple bioinformatics grid. *BMC Bioinformatics* 6, 197. doi: 10.1186/1471-2105-6-197.

Carver, T., and Bleasby, A. (2003). The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics* 19, 1837–1843.

Cerami, E. G., Bader, G. D., Gross, B. E., and Sander, C. (2006). cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics* 7, 497. doi: 10.1186/1471-2105-7-497.

Cheung, K.-H., Yip, K. Y., Smith, A., deKnikker, R., Masiar, A., and Gerstein, M. (2005). YeastHub: a semantic Web use case for integrating data in the life sciences domain. *Bioinformatics* 21, i85–i96.

CIDMS (2007). *Cardiac Integrated Database Management System for Cardiac Systems Biology*. Available at: www.icsb-2007.org/proceedings/abstracts/G06.pdf.

CIDMS-PD (2009). *An Innovative Systems Biology Knowledgebase for Cellular Pathways*. Available at: http://cidms.org/pathways.

Clark, T., Jurek, J., Kettler, G., and Preuss, D. (2005). A structured interface to the object-oriented genomics unified schema for XML-formatted data. *Appl. Bioinformatics* 4, 13–24.

Ćurčin, V., Ghanem, M., Guo, Y., Köhler, M., Rowe, A., Syed, J., and Wendel, P. (2002). "Discovery net: toward a grid of knowledge discovery," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Edmonton.

Czauderna, T., Klukas, C., and Schreiber, F. (2010). Editing, validating, and translating of SBGN maps. *Bioinformatics* 26, 2340–2341.

Deelman, E., Singh, G., Su, M. H., Blythe, J., Gil, Y., Kesselman, C., Mehta, G., Vahi, K., Berriman, G. B., Good, J., Laity, A., Jacob, J. C., and Katz, D. S. (2005). Pegasus: a framework for mapping complex scientific workflows onto distributed systems. *Sci. Program.* 13, 219–237.

Fenner, J. W., Brook, B., Clapworthy, G., Coveney, P. V., Feipel, V., Gregersen, H., Hose, D. R., Kohl, P., Lawford, P., McCormack, K. M., Pinney, D., Thomas, S. R., Van Sint Jan, S., Waters, S., and Viceconti, M. (2008). The EuroPhysiome, STEP and a roadmap for the virtual physiological human. *Philos. Transact. R. Soc. A Math. Phys. Eng. Sci.* 366, 2979–2999.

Finney, A., and Hucka, M. (2003). Systems biology markup language: level 2 and beyond. *Biochem. Soc. Trans.* 31, 1472–1473.

Foster, I. (2005). Service-oriented science. *Science* 308, 814–817.

Funahashi, A., Jouraku, A., Matsuoka, Y., and Kitano, H. (2007). Integration of CellDesigner and SABIO-RK. *In Silico Biol.* 7, S81–S90.

Gehlenborg, N., O'Donoghue, S. I., Baliga, N. S., Goesmann, A., Hibbs, M. A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., and Gavin, A. C. (2010). Visualization of omics data for systems biology. *Nat. Methods* 7, S56–S68.

Gentleman, R. C., O'Donoghue, S. I., Baliga, N. S., Goesmann, A., Hibbs, M. A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., and Gavin, A. C. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.01–R80.16.

Goble, C. A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P., and De Roure, D. (2010). MyExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucl. Acids Res.* 38, W677–W682.

Goble, C. A., and De Roure, D. (2008). Curating Scientific Web Services and Workflows. *Educause* 43, 1527–6619.

Good, B. M., and Wilkinson, M. D. (2006). The life sciences semantic web is full of creeps!. *Brief. Bioinform.* 7, 275–286.

Gordon, P. M., Trinh, Q., and Sensen, C. W. (2007). Semantic web service provision: a realistic framework for bioinformatics programmers. *Bioinformatics* 23, 1178–1180.

Guarino, N. (1998). "Formal ontology in information systems," in *Proceedings of the 1st International Conference June 6–8*, 1998. Trento: IOS Press.

Gudivada, R. (2007). "A genome–phenome integrated approach for mining disease-causal genes using Semantic Web," in *Health Care and Life Sciences*

Data Integration for the Semantic Web, Sixteenth International World Wide Web Conference (WWW2007) Workshops, Athens, GA.

Haas, L. M., Schwarz, P. M., Kodali, P., Kotlar, E., Rice, J. E., and Swope, W. C. (2001). DiscoveryLink: a system for integrated access to life sciences data sources. *IBM Syst. J.* 40, 489–511.

Haider, S., Ballester, B., Smedley, D., Zhang, J., Rice, P., and Kasprzyk, A. (2009). BioMart Central Portal – unified access to biological data. *Nucl. Acids Res.* 37, W23–W27.

Harris, P. J., Buyya, R., Chu, X., Kobialka, T., Kazmierczak, E., Moss, R., Appelbe, W., Hunter, P. J., and Thomas, S. R. (2009). The Virtual Kidney: an eScience interface and grid portal. *Philos. Transact. R. Soc. A: Math. Phys. Eng. Sci.* 367, 2141–2159.

Harris, S. and Gibbins, N. (2003). *3Store: Efficient Bulk RDF Storage. PSSS*. CEUR-WS.org.

Hey, T., and Trefethen, A. (2003). e-Science and its implications. *Philos. Transact. R. Soc. Lond, A: Math. Phys. Eng. Sci.* 361, 1809–1825.

Hmida, M. M. B., Tomaz, R. F., and Monfort, V. (2005). "Applying AOP concepts to increase web services flexibility," in *Proceedings of the International Conference on Next Generation Web Services Practices*. IEEE Computer Society.

Hunter, P., Coveney, P. V., de Bono, B., Diaz, V., Fenner, J., Frangi, A. F., Harris, P., Hose, R., Kohl, P., Lawford, P., McCormack, K., Mendes, M., Omholt, S., Quarteroni, A., SkÃ¥r, J., Tegner, J., Randall Thomas, S., Tollis, I., Tsamardinos, I., van Beek, J. H. G. M., and Viceconti, M. (2010). A vision and strategy for the virtual physiological human in 2010 and beyond. *Philos. Transact. R. Soc. A: Math. Phys. Eng. Sci.* 368, 2595–2614.

Hunter, P. J., Crampin, E. J., and Nielsen, P. M. F. (2008). Bioinformatics, multiscale modeling and the IUPS physiome project. *Brief. Bioinform.* 9, 333–343.

Huss, J. W., 3rd, Orozco, C., Goodale, J., Wu, C., Batalov, S., Vickers, T. J., Valafar, F., and Su, A. I. (2008). A gene wiki for community annotation of gene function. *PLoS Biol.* 6, e175. doi: 10.1371/journal.pbio.0060175.

Jenkinson, A., Albrecht, M., Birney, E., Blankenburg, H., Down, T., Finn, R., Hermjakob, H., Hubbard, T., Jimenez, R., Jones, P., Kahari, A., Kulesha, E., Macias, J., Reeves, G., and Prlic, A. (2008). Integrating biological data – the distributed annotation system. *BMC Bioinformatics* 9, S3. doi: 10.1186/1471-2105-9-S8-S3.

Jurisica, I., Mylopoulos, J., and Yu, E. (2004). Ontologies for knowledge management: an information systems perspective. *Knowl. Inf. Syst.* 6, 380–401.

Kamel Boulos, M. N., and Wheeler, S. (2007). The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. *Health Info. Libr. J.* 24, 2–23.

Kawas, E., Senger, M., and Wilkinson, M. D. (2006). BioMoby extensions to the Taverna workflow management and enactment software. *BMC Bioinformatics* 7, 523. doi: 10.1186/1471-2105-7-523.

Kiczales, G., Lamping, J., and Mendhekar, A. (1997). "Aspect oriented programming," in *The 11th European Conference of Object-Oriented Programming*, LNCS 1241, Jyväskylä.

Kitano, H., Funahashi, A., Matsuoka, Y., and Oda, K. (2005). Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* 23, 961–966.

Komatsoulis, G. A., Warzel, D. B., Hartel, F. W., Shanbhag, K., Chilukuri, R., Fragoso, G., Coronado, S., Reeves, D. M., Hadfield, J. B., Ludet, C., and Covitz, P. A. (2008). caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability. *J. Biomed. Inform.* 41, 106–123.

Krause, F., Uhlendorf, J., Lubitz, T., Schulz, M., Klipp, E., and Liebermeister, W. (2010). Annotation and merging of SBML models with semanticSBML. *Bioinformatics* 26, 421–422.

Kushida, T., Takagi, T., and Fukuda, K. I. (2006). Event ontology: a pathway-centric ontology for biological processes. *Pac. Symp. Biocomput.* 11, 152–163.

Laibe, C., and Le Novere, N. (2007). MIRIAM Resources: tools to generate and resolve robust cross-references in systems biology. *BMC Syst. Biol.* 1, 58. doi: 10.1186/1752-0509-1-58.

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. (2007). AlzPharm: integration of neurodegeneration data using RDF. *BMC Bioinformatics* 8(Suppl. 3), S4. doi: 10.1186/1471-2105-8-S3-S4.

Langmead, B., Schatz, M. C., Lin, J., Pop, M., and Salzberg, S. L. (2009). Searching for SNPs with cloud computing. *Genome Biol.* 10, R134.

Le Novere, N. (2006). Model storage, exchange and integration. *BMC Neurosci.* 7(Suppl. 1), S11. doi: 10.1186/1471-2202-7-S1-S11.

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, Tom C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., and Kitano, H. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741.

Lee, T. J., Pouliot, Y., Wagner, V., Gupta, P., Stringer-Calvert, D. W. J., Tenenbaum, J. D., and Karp, P. D. (2006). BioWarehouse: a bioinformatics database warehouse toolkit. *BMC Bioinformatics* 7, 170. doi: 10.1186/1471-2105-7-170.

Li, C., Courtot, M., Le Novere, N., and Laibe, C. (2009). BioModels.net web services, a free and integrated toolkit for computational modelling software. *Brief. Bioinform.* 11, 270–277.

Lloyd, C. M., Halstead, M. D., and Nielsen, P. F. (2004). CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* 85, 433–450.

Ludäscher, B., Ludscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E. A, Tao, J., and Zhao, Y. (2006). Scientific workflow management and the Kepler system. *Concurr. Comput.* 18, 1039–1065.

Maier, C. W., Long, J. G., Hemminger, B. M., and Giddings, M. C. (2009). UltraStructure database design methodology for managing systems biology data and analyses. *BMC Bioinformatics* 10, 254. doi: 10.1186/1471-2105-10-254.

Martin, A. C. R. (2001). Can we integrate bioinformatics data on the internet? *Trends Biotechnol.* 19, 327–328.

Martin, S., Hohman, M. M., and Liefeld, T. (2005). The impact of life science identifier on informatics data. *Drug Discov. Today* 10, 1566–1572.

Matsuoka, Y., Ghosh, S., Kikuchi, N., and Kitano, H. (2010). Payao: a community platform for SBML pathway model curation. *Bioinformatics* 26, 1381–1383.

Mazza, T. (2010). Editorial: accelerating systems biology. *Brief. Bioinform.* 11, 267–269.

McIlraith, S. A., Son, T. C., and Honglei, Z. (2001). Semantic web services. *Intell. Syst., IEEE* 16, 46–53.

Miller, J., Kunal, V., Rajasekaran, P., Sheth, S., Agggarwal, R., and Sivashanmugam, K. (2004). *WSDL-S: Adding Semantics to WSDL – White Paper*. Available at: http://lsdis.cs.uga.edu/library/download/wsdl-s.pdf

Mlecnik, B., Scheideler, M., Hackl, H., Hartler, J., Sanchez-Cabo, F., and Trajanoski, Z. (2005). PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucl. Acids Res.* 33, W633–W637.

Mons, B., Ashburner, M., Chichester, C., van Mulligen, E., Weeber, M., den Dunnen, J., van Ommen, G. J., Musen, M., Cockerill, M., Hermjakob, H., Mons, A., Packer, A., Pacheco, R., Lewis, S., Berkeley, A., Melton, W., Barris, N., Wales, J., Meijssen, G., Moeller, E., Roes, P. J., Borner, K., and Bairoch, A. (2008). Calling on a million minds for community annotation in WikiProteins. *Genome Biol.* 9, R89.

Neerincx, P. B. and Leunissen, J. A. (2005). Evolution of web services in bioinformatics. *Brief. Bioinform.* 6, 178–188.

Nord, R. L., and Czarnecki, K. (2004). *Generative Software Development*. Springer Berlin/Heidelberg: Software Product Lines, 148–151.

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., and Musen, M.A. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucl. Acids Res.* gkp440, 37.

Oinn, T., Greenwood, M., Addis, M., Alpdemir, M. N., Ferris, J., Glover, K., Goble, C., Goderis, A., Hull, D., Marvin, D., Li, P., Lord, P., Pocock, M.R., Senger, M., Stevens, R., Wipat, A., and Wroe, C. (2006). Taverna: lessons in creating a workflow environment for the life sciences. *Concurr. Comput.* 18, 1067–1100.

Pico, A. R., Kelder, T., van Iersel, M. P., Hanspers, K., Conklin, B. R., and Evelo, C. (2008). WikiPathways: pathway editing for the people. *PLoS Biol.* 6, e184. doi: 10.1371/journal.pbio.0060184.

Post, L. J. G., Roos, M., Marshall, M. S., van Driel, R., and Breit, T. M. (2007). A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data. *Bioinformatics* 23, 3080–3087.

Puustjarvi, J., and Puustjarvi, L. (2009). The role of medicinal ontologies in querying and exchanging pharmaceutical information. *Int. J. Electron. Healthc.* 5, 1–13.

Reddy, M. P., Prasad, B. E., and Reddy, P. G. (1994). A methodology for integration of heterogeneous databases. *IEEE Trans. Knowl. Data Eng.* 6, 920–933.

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., and Cheung, K. H. (2007). Advancing translational

research with the Semantic Web. *BMC Bioinformatics* 8(Suppl. 3), S2. doi: 10.1186/1471-2105-8-S3-S2.

Ruttenberg, A., Rees, J. A., Samwald, M., and Marshall, M. S. (2009). Life sciences on the semantic web: the Neurocommons and beyond. *Brief. Bioinform.* 10, 193–204.

Safran, M., Solomon, I., Shmueli, O., Lapidot, M., Shen-Orr, S., Adato, A., Ben-Dor, U., Esterman, N., Rosen, N., Peter, I., Olender, T., Chalifa-Caspi, V., and Lancet, D. (2002). GeneCards 2002: towards a complete, object-oriented, human gene compendium. *Bioinformatics* 18, 1542–1543.

Sagotsky, J. A., Zhang, L., Wang, Z., Martin, S., and Deisboeck, T. S. (2008). Life sciences and the web: a new era for collaboration. *Mol. Syst. Biol.* 4, 201.

Saltz, J., Oster, S., Hastings, S., Langella, S., Kurc, T., Sanchez, W., Kher, M., Manisundaram, A., Shanbhag, K., and Covitz, P. (2006). caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics* 22, 1910–1916.

Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Lu, Z., Madden, T. L., Madej, T., Maglott, D. R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Wang, Y., John Wilbur, W., Yaschenko, E., and Ye, J. (2010). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 38, D5–D16.

Schatz, M. C., Langmead, B., and Salzberg, S. L (2010). Cloud computing and the DNA data race. *Nat Biotechnol.* 28, 691–693.

Shah, S. P., Huang, Y., Xu, T., Yuen, M. M., Ling, J., and Ouellette, B. F. (2005). Atlas – a data warehouse for integrative bioinformatics. *BMC Bioinformatics* 6, 34. doi: 10.1186/1471-2105-6-34.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.

Sheth, A. P., and Larson, J. A. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv.* 22, 183–236.

Splendiani, A. (2008). RDFScape: semantic web meets systems biology. *BMC Bioinformatics* 9(Suppl. 4), S6. doi: 10.1186/1471-2105-9-S4-S6.

Stein, L. (2010). The case for cloud computing in genome informatics. *Genome Biol.* 11, 207.

Stein, L. D. (2003). Integrating biological databases. *Nat. Rev. Genet.* 4, 337–345.

Stevens, R., Aranguren, M. E., Wolstencroft, K., Sattler, U., Drummond, N., Horridge, M., and Rector, A. (2007). Using OWL to model biological knowledge. *Int. J. Hum. Comput. Stud.* 65, 583–594.

Stevens, R., Paton, N. W., Baker, P., Ng, G., Goble, C. A., Bechhofer, S., and Brass, A. (1999). "TAMBIS Online: a bioinformatics source integration tool," in *Proceedings of the 11th International Conference on Scientific and Statistical Database Management*. IEEE Computer Society.–

Tan, W., Missier, P., Foster, I., Madduri, R., De Roure, D., and Goble, C. (2010). A comparison of using Taverna and BPEL in building scientific workflows: the case of caGrid. *Concurr. Comput.* 22, 1098–1117.

Taylor, I. J., Deelman, E., Gannon, D. B., Shields, M., Taylor, I., Wang, I., and Harrison, A. (2007). *The Triana Workflow Environment: Architecture and Applications*. London: Springer, In Workflows for e-Science, 320–339.

Tyrelle, G. (2005). *The Semantic Web for Life Sciences Now*! Available at: http://archive.nodalpoint.org/node/1704

Van Hemert, J. L., and Dickerson, J. A. (2010). PathwayAccess: celldesigner plugins for pathway databases. *Bioinformatics.*

W3C. *W3C.* (2002). *Web Services Activity group*. Available at: http://www.w3.org/2002/ws/

W3C/CWI. (2010). *XForms*. Available at: http://www.w3.org/MarkUp/Forms/

Walton, N. A., Brenton, J. D., Caldas, C., Irwin, M. J., Akram, A., Gonzalez-Solares, E., Lewis, J. R., Maccallum, P. H., Morris, L. J., and Rixon, G. T. (2010). PathGrid: a service-orientated architecture for microscopy image analysis. *Philos. Transact. R. Soc. A Math. Phys. Eng. Sci.* 368, 3937–3952.

Wang, X., Gorlitsky, R., and Almeida, J. S. (2005). From XML to RDF: how semantic web technologies will change the design of "omic" standards. *Nat. Biotechnol.* 23, 1099–1103.

Wang, Z., Gao, X., He, C., Miller, J. A., Kissinger, J. C., Heiges, M., Aurrecoechea, C., Kraemer, E.T., and Pennington, C. (2007). "A comparison of federated databases with web services for the integration of bioinformatics data," in *Conference on Bioinformatics and Computational Biology (BIOCOMP)*, Vegas, NV.

WikiPathways (2008). GPML plugin for Cytoscape. Available on: http://www.pathvisio.org/wiki/Cytoscape_plugin.

Wilkinson, M., and Links, M. (2002). BioMOBY: an open source biological Web services proposal. *Brief. Bioinformatics* 3, 331–341.

Wilkinson, M., Schoof, H., Ernst, R., and Haase, D. (2005). BioMOBY successfully integrates distributed heterogeneous bioinformatics web services. The PlaNet exemplar case. *Plant Physiol.* 138, 5–17.

Wood, J. D., (2004). The first nobel prize for integrated systems physiology: Ivan Petrovich Pavlov, 1904. *Physiology* 19, 326–330.

Zdobnov, E. M., Lopez, R., Apweiler, R., and Etzold, T. (2002). The EBI SRS server – recent developments. *Bioinformatics* 18, 368–373.

Zhang, Z., Cheung, K.-H., and Townsend, J. P. (2009). Bringing web 2.0 to bioinformatics. *Brief. Bioinform.* 10, 1–10.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 August 2010; paper pending published: 28 September 2010; accepted: 18 October 2010; published online: 03 December 2010.

Citation: Sreenivasaiah PK and Kim DH (2010) Current trends and new challenges of databases and web applications for systems driven biological research. Front. Physio. 1:147. doi: 10.3389/fphys.2010.00147

This article was submitted to Frontiers in Systems Physiology, a specialty of Frontiers in Physiology.