Technical Perspectives on Knowledge Management in Bioinformatics Workflow Systems

Walaa N. Ismail
Information Systems Dept.
Faculty of Computers and Information
King Saud University.
Saudi Arabia

M.Sabih Aksoy
Information Systems Dept.
Faculty of Computers and Information
King Saud University.
Saudi Arabia

Abstract—**Workflow systems by it's nature can help bioinformaticians to plan for their experiments, store, capture and analysis of the runtime generated data. On the other hand, the life science research usually produces new knowledge at an increasing speed; Knowledge such as papers, databases and other systems knowledge that a researcher needs to deal with is actually a complex task that needs much of efforts and time. Thus the management of knowledge is therefore an important issue for life scientists. Approaches has been developed to organize biological knowledge sources and to record provenance knowledge of an experiment into a readily resource are presently being carried out. This article focuses on the knowledge management of in silico experimentation in bioinformatics workflow systems.**

Index Terms—**Bioinformatics Workflows; Knowledge Management; Ontologies; Scientific Workflows; Semantic Web.**

I. INTRODUCTION

Scientific experimentation in science domains contains all required aspects of the experimentation process including data analysis, modeling, and testing [1].

A workflow is a well-defined organization of activities or patterns designed to achieve a certain data transformation [2]. Workflow systems by it's nature can help bioinformaticians to plan for their experiments, store, capture and analyses of the runtime generated data [2]. Complex workflow systems that integrate programs, methods, agents, and services coming from diverse organizations or sites requires a more flexible framework that can execute such complex scenario [3] . In such a way, the execution sequence and the scheduling of algorithms, data, services, and other software components are orchestrated in a single virtual framework [3].

Workflow systems construction process include but not limited to the following steps [2]:

(a) Users define typical execution patterns for computational process.
(b) The system store the generated pattern form step 1.
(c) Later the users can retrieve such pattern for modifications and re-execute them in different scenarios.

Figure 1 shows a simple workflow example for constructing phylogenetic tree to a given sequence. The steps required to construct phylogenetic tree [4] which begin with the user input sequence query is:

Figure 1: A Simple Workflow For Constructing Phylogenetic Tree

Step 1   Choose an appropriate markers for the phylogenetic analysis from the workflow database.
Step 2   Perform multiple sequence alignments for the matched sequences in step1.
Step 3   Select an evolutionary model.
Step 4   Reconstruct the Phylogenetic tree.
Step 5   Evaluate the phylogenetic tree.

Consider for example **Step 2**; a wide range of algorithms can perform the alignment like
(a) synchronous Blast services, and (b) Blast services. Hence, the workflow enable the user simply specifies that *a Sequence Alignment* is desired. On the second hand expert users can choose to specify all the analysis required for every step in the workflow.

Our level of understanding will be increased to more effectively solve problems and make required decision via knowledge management (KM) discipline [5]. KM is subject that provides strategy, process and technology to share information and expertise among users.

KM has been an important subject disciplines for many fields today which needs understanding knowledge processes and selecting the most appropriate KM systems that can help in creating, storing, and more effectively sharing knowledge [5].

A bioinformatics workflow system seems by its nature could benefit from KM principles and methodology [6] [7]. The main reasons are bioinformatics workflow systems usually interact with [8]:

1) A modern infrastructure that are frequently change.
2) General community knowledge.
3) The biologists who generally prefer to share their knowledge with each others.
4) Also those workflows usually acquire fast accessible knowledge sources.

Therefore, in the bioinformatics, KM can be defined as a systematic process that allow creating, capturing, sharing, and analyzing knowledge in ways that affect system performance and availability [9].

With the vast amount of the available bioinformatics tools, services and algorithms that can execute the biologists tasks; it's a must to have certain technology that allow automation and discovery of such resources, in addition to that the bioinformaticians need to create complex workflows from a wide range of available web services knowledge base. So far the emergence of the semantic Web technology (SW) [10] is starting to have a significant impact on knowledge integration, querying, and knowledge sharing in the life science domain [10], [11].
The success of knowledge management system (KMS) in Bioinformatics can be achieved by the assistance of knowledge technology. Knowledge technology is a part of KM, refers to an unclear set of available tools that enable better representation, organization and exchange of information and knowledge [8] [1]. Among the existence technologies are knowledge mapping, collaborative technologies, semantic technologies and social computing tools [12].

The growing acceptance of the semantic web as a means to manage biological knowledge is noteworthy [1] as SW technology offers more flexibility in data modeling by integration of large amounts of data [11]. Therefore, This paper will discuss the technical perspectives on KMS in Bioinformatics that focus on technology, ideally those that enhance knowledge sharing and growth in bioinformatics domain.

The rest of the paper is organized as follows: Section II, presents how semantic Web technology is an effective knowledge management technology in life science domain. Section III, discussed knowledge management efforts in Bioinformatics workflow systems and presents the knowledge management life cycle in bioinformatics workflow systems. Section IV explain semantic system biology cycle. Section V presents related work about workflow and workflow systems in life science. Finally, section VI concludes and outlines directions for future work.

II. Towards effective knowledge management in the Life Sciences

Semantic Web (SW) technology is an effective knowledge management technology in life science, since it allow automatic discovery and execution of web services that can handle the workflow tasks, which prevents biologists from the need to working with similar or time-consuming tasks, such as taking manual copy of one tool and then pasting that tool to another tool [10].

SW depends on a set of web technologies specifically designed to facilitate automated machine interoperability [10]. It promises to meet the challenge of integrating and querying highly diverse and distributed resources [13].

Systems based on SW would provide sophisticated frameworks to manage and retrieve knowledge. Ontologies in biology (bio-ontologies) and the semantic Web are playing a vital role in the integration of data and knowledge by offering an explicit, unambiguous and rich data and knowledge representation mechanisms [14] [10].

Biomedical ontologies are playing an important role in life sciences semantic web since they help in capturing the semantics of entities and their interrelationships within biology domain, thereby reducing conceptual ambiguity, increasing reusability and computational automation that aids in knowledge gathering and discovery [15].
Ontologies can be classified according to the degree of conceptualization which includes [12]:

1) **Upper-level ontology:** Ontologies that describes general concepts which are independent of a particular domain. Their applicability is in providing support to a large number of ontologies. The Basic Formal Ontology[1] is a widely used upper level ontology in a number of subdomains within the life sciences.

2) **Domain ontology:** The knowledge represented in this type of ontology serves a particular domain by providing vocabularies about concepts and their relationships governing the domain such as The Gene Ontology (GO)[2].

3) **Application ontology:** These ontologies are typically used to define concepts for a particular use case. For instance, EFO[3] is used to represent concepts and sample variables from gene expression experiments. An ontology that captures knowledge related to the cell cycle processes.

III. Workflows at the Knowledge Level

Bioinformatics workflow systems could benefit from KM efforts that define strategies to capture the vast amount of available bioinformatics tools, services and algorithms that can execute a certain biologists tasks. Knowledge management (KM) processes encompasses many tasks such as knowledge formulation, storage and distribution [14]. Figure 2 represent knowledge management life cycle in bioinformatics [14]

1) The **Creation stage** identify the major bioinformatics system components including rich knowledge base about services/tools , algorithms and data conversion methods.

[1]http://www.ifomis.org/bfo/
[2]http://geneontology.org/
[3]http://bioportal.bioontology.org/ontologies/EFO

2) The **Identify or collect** stage collects the local and shared knowledge, algorithms, other workflows provenance, scientific theories and available scientists experience to create the selected main knowledge domains components.
3) The **Select** stage takes the composed collected knowledge and evaluate its value. For organizing and classifying knowledge that will be stored in the knowledge repositories; one framework should be selected..
4) The **Store** stage classifies the collected knowledge and adds them to the workflow system.
5) The **Share** stage retrieves knowledge from the workflow system and makes it available to the system users. Scientists often needs to share and use ideas, results of experiments, knowledge expertise over the network or from other workflow systems.
6) The **Apply** stage reuses the collected knowledge in executing workflow tasks, building new workflow, discovering new research ideas, taking important decisions and learning new thought.
7) The **Update** stage provides a creative update and automated knowledge discovery platform by investigating uncovered new knowledge, such as new methods, algorithms, scientists feedback, analysis, research, and experimentation.
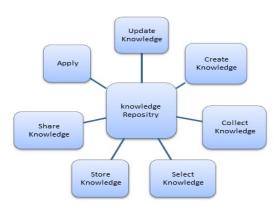

Figure 2: Knowledge Management Life Cycle

Workflow systems can be defined as repositories of scientific knowledge [2] [16]; so Does describing workflow systems at the knowledge level could define new concepts? if so, we have to ask what should workflow systems expected to achieve by using that knowledge?

Figure 3 shows a set of layers in the workflows specification process, the layers organized such that from more abstraction level to more specific level. The information contained on each layer can be used to implement the layer below it. Workflows specify what data will be used as well as the services or codes that are to be used to execute each workflow task. Those refers to layers 2 and 1.

The data and services that has been specified in *level 1* workflows are then mapped to actual execution resources in the execution environment, resulting in level 0 workflows.

Moving up in the figure levels, some workflow steps can be ignored if they are not central to the experiment a workflow can then be described not by the specific resource but by identifying classes of services to be used instead.

For example, if a workflow to Construct phylogenetic tree [4] is needed; user query sequence is first processed with normalization step followed by sequence alignment step, followed by selection of an evolutionary model, and then Phylogenetic tree reconstruction without specifying any algorithm or methods to be used.

Workflows at level 3 does not specify how each operation will be executed in relation to other operation in the workflow instead it specify how data will be carried out. At a highest level of workflow abstraction, only the desired results would be specified without any other details. For example, Construct phylogenetic tree to a reference sequence without any details are provided about how to construct the tree or what workflow to be used or what type of data to be generated.

Having scientific workflow means to have a wide range of methods, algorithms, and tools that can perform a given workflow task at different level of granularity; in addition to that, the architecture at *the symbol level* describes the capabilities of workflow systems and how to execute the workflow identified tasks [2].

On the other hand, *the knowledge level* describe the scientific tasks that a workflow system expected to accomplish through suggestion of descriptions and capabilities that would affect what can be done. With more knowledge about workflows usage and integration will improve the workflow behavior by solving more tasks and producing new kinds of results [2], [17].

Figure 3, also relate workflow abstraction layers to the knowledge level and the symbol level. In summary having systems that can take workflows requests from users, and then execute the workflow without any details about execution details or resources would decrease the inexperienced user overload who have small amount of knowledge about the workflows tasks selection and execution.

IV. REASONING WITH WORKFLOWS AT THE KNOWLEDGE LEVEL

To receive the accurate knowledge that can improve any system performance requires a system that can determines the user purposes, and then tracks the user's actions and behavior. [18]
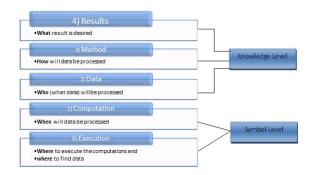
Figure 3: Workflow Abstraction Layers

Semantic system biology (SSB) provides a semantic description of the knowledge about the biological systems on the whole facilitating data integration, knowledge management, reasoning and querying [7]. Figure 4 describe
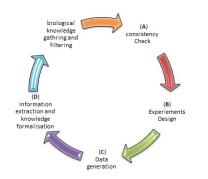

Figure 4: Semantic System Biology Cycle

the semantic systems biology cycle [10]. The cycle begins with gathering and integrating biological knowledge into a semantic knowledge base; then data are checked for consistency, then; (B)This yields criteria about particular functions of biological components that may be used to design experiments; (C) The experiments generates new data and might also verify the design criteria.(D)The new data are then integrated into the knowledge base, thereby enhancing the quality of the knowledge base and allowing a new cycle of hypothesis building and experimentation.

With knowledge of what workflow components do, and experiment design; workflow systems can assist scientists by using those knowledge to make automatically decisions concerned about specific domain [19].

Figure 5 is a schematic representation of the workflow in SSB [13]. Firstly, biological knowledge is extracted from disparate resources and integrated into a knowledge base.

Given the user query and knowledge base about the application domain and input data: the reasoner identifies the strategies to the user to use and run the tools that can execute

his/her request. Executor then run the completed workflow and updates the knowledge base with the results of workflow execution that can be used to make new inference.
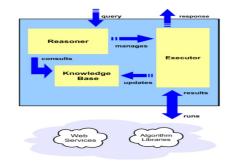

Figure 5: Reasoning in Workflow system at knowledge Level

V. WORKFLOWS AND WORKFLOW SYSTEMS

There have been a large number of workflow design and execution engines that supports In silico biological study. Workflows Systems have been used as a useful paradigm to model and manage complex scientific analyses [2] [3]. Some scientific workflows represent how to use and compose a variety of remote services to accomplish an overall task. Other workflows systems have incorporated semantic component to support knowledge representation of workflows execution.

Galaxy in [15] is a workflow web-based platform for analysing genomic sequences. In Galaxy several tools can be merged, ranging from simply manipulating data to complex analysis tasks. Galaxy provide an flexible construction of workflows as it can:

- Combine knowledge of current workflow tasks.
- Can be executed from a single Web interface.
- Share the output of the tool by sending the current results to other tools as input.
- Store the history of all performed actions which facilitate the analyses of any task at any time.
- Galaxy can use users history to build workflow.

- The workflows can be re-used in other systems, like other servers or myExperiment [15].
- The capturing of data provenance and the context of a workflow are automatically tracked and managed.

Wings [20] a workflow system that allow users to describe their desired analyses tasks. After the users describe their goal Wings begin automatically to validate the input goal and data by using a knowledge base (using ontologies and rules) about workflow components and finally map each task to services that Pegasus [16] use to execute that task.

Wings organize all workflow components in hierarchies; components such as workflow tasks, data, properties, and constraints regarding their proper use. In addition Wings allow users to describe a workflow templates that can be reused for different scenarios, and it also can automatically build workflows using data products descriptions of what the user prefer.

In [21], the author presented Sesame a semantic bioinformatics workflow design system with new ontology for bioinformatics tools/services. Sesame allows the biologists to perform their analyses using terms that they are familiar with. After designing the semantic workflow, Sesame have a knowledge repository that associating each analyses entity with the instances of bioinformatics tools/services and data that previously had been used to handle such data and tasks. Sesame free the biologists from the necessity of learning the details of the computational aspects of the bioinformatics tools. Also, Sesame can perform simple instantiation cases and for each analyses entity Sesame ask the user to select one instance of bioinformatics tools/services. Then, the user specifies the parameters and input data for the selected tool.

Taverna [22] is a workflow building platform that facilitate the matching process of users requests with the available workflows templates and services via using of rich knowledge descriptions of workflows components that enable users to specify either the type of service they wish to use or a graph of workflow services and their dataflow. On the other hand Taverna is designed as a do-it-all platform, which can be very complex to be used for biologists with limited computing background.

The authors in [13] have utilized several semantic technologies to identify the scientists intent, and then to facilitate the control of workflow execution and enrichment of workflow provenance of new tasks,

The Magallanes [23] is a library that can develop effective workflow discovery engines that can help to collect web-services which will be used to execute workflow tasks and it's datatypes. The discovery of Web services can be based on *syntax description of services or objects* that is it's name which is often unsatisfactory in bioinformatics because it presumes knowledge of objects names or semantics services discovery by specifying a general descriptions about services or objects which allow to have a more accurate discovery mechanism. Magallanes uses a syntactic text-based approach and a semantic approach to collect different services that can handle the input and output data types.

In [24] a framework for services selection in the life-sciences is proposed. The solution build workflows by data-type matching methods that provide less time and effort through selection of best services that can handle workflow tasks so that a small set of the available services that can achieve user task are identified.

Kepler [25] is a graphical system for scientific workflows design, execute, reuse, and sharing. Kepler's provide high effective workflow designing process by monitoring data and provenance information during the initial workflow design stage Kepler supports also many advanced features such as automatic workflow validation and editing; by providing a semantic annotation of workflow tasks from a domain ontology. Also Kepler's workflows are created by connecting a chain of workflow components together called *Actors* each Actor has several ports through which input and output ports containing data and data references are sent and received. Each workflow has a Director that determines the model of computation used by the workflow,

The knowledge level of any intelligent workflow system is concerned with the kind of knowledge it can use, and how it response to users requests, or what is the user's goals. [1]

The initiatives proposed comparisons given in Table I demonstrate what is the advantages of the semantic web technologies workflow design systems [10], including;

- **Automatic workflow generation.** During the building of workflows the system can automatically build the workflow without the need to any other tools as it has it's own knowledge about the workflow components, and data.
- **Workflow validation ;** the knowledge of components and data that the system have about the different operations enables the workflow system to validate any workflow task even in a complex composition scenarios.
- **Automatic metadata generation;** the descriptions of new data products that are generated during workflow execution are automatically generated as we have knowledge-level descriptions on each datasets.
- **Guarantee of trusted provenance;** Learning from previous designs that perform a similar task when designing a new workflow is more economic and efficient. That is; if the provenance of new data products obtained through a highly efficient presumed workflow systems it can be a good indicator of high quality process used to obtain those new results.

The system can include knowledge base that integrate its components with **Semantic about provenance** that comprises the experiment with all the other metadata about experiments which help the scientist to learn how to use and compose

Table I: Workflow Systems at Knowledge Level

Feature examined	Galaxy	Wings	Taverna	Sesame	Kepler	Magallanes
Can workflow generated automatically?	Yes	Yes	Yes	Yes	Yes	Yes
Does the user queries is validated?	Yes	Yes	Yes	Yes	Yes	Yes
Can workflows be shared with other users?	Yes	Yes	Yes	Yes	Yes	No
Can results be shared with other users?	Yes	Yes	Yes	Yes	Yes	Yes.
Data provenance available?	Yes	Yes	No	Yes	Yes	No
Can a user add a web service to the tool?	No	Yes	Yes	Yes	No	Yes
The description of new data products?	Limited	Yes	Yes	Yes	Yes	Yes
Contain widely accepted workflows?	No	Yes	Yes	Yes	Yes	No
Automation of workflow execution.	Yes	Yes	Yes	Yes	Yes	Yes

services in a another workflow systems. Also mining provenance data of repeatedly executed workflow tasks could help to identify the performance and quality information about those services that can execute a similar tasks or accepting the same inputs data type. This information can assist the scientist to choose between vast amount of alternative services. [7] [14]

VI. CONCLUSIONS AND FUTURE WORK

Knowledge management is a broadly defined concept varying from one domain to the other. For instance, knowledge management in the business domain [26] would mainly deal with management of business activities such as business policies, assets and risk assessments. In comparison, knowledge management in bioinformatics [19] deals with management of what is understood about the various components of a system of interest. Also knowledge representation plays a crucial role in the facilitation of processing and sharing knowledge between people and application systems.

Additionally, knowledge representation languages should adopt a common syntax that is reusable and enables parsing of data in a semantically unambiguous manner [5]. Ontologies in biology (bio-ontologies) and the semantic Web are playing vital role in the integration of data and knowledge by offering an explicit, unambiguous and rich representation mechanism. This increased influence led to the proposal of the semantic systems biology paradigm to complement the techniques currently used in systems biology. semantic systems biology provides a semantic description of the knowledge about the biological systems on the whole facilitating data integration, knowledge management, reasoning and querying. These conditions in scientific workflow environment will support intelligent inferencing of facts over a given biological domain and also facilitate processing of information even in complex scenarios that require composition of different sources, or algorithms to be handled.

For future work workflow system could benefit from identifying syntactic patterns [27] which are sets of axioms in an OWL ontology with a regular structure. Detecting these patterns and reporting them in human readable form should help the inexperienced workflow users to understand the style of ontology and is therefore useful in expressing the bioinformatics experiments knowledge more preciously. However, the detection of such patterns is sensitive to variations in the assertions [27].

Also its a must to differentiate between axioms that are semantically equivalent but syntactically different as in this case it can lead to reducing the effectiveness of the knowledge

presented in any workflow system [18]. So workflow methods could focuses on Semantic regularity analysis that focuses on the knowledge encoded in the ontology, rather than how it is spelled.

REFERENCES

[1] P. Zhang, L. Zhang, Z. Fan, and X. Qiu, "Knowledge based modeling method of artificial society oriented to emergency management." Springer Berlin Heidelberg, 2014, vol. 461, pp. 278–287.

[2] P. Walsh, J. Carroll, and R. D. Sleator, "Accelerating in silico research with workflows: A lesson in simplicity," *Computers in Biology and Medicine*.

[3] S. Y. Bendoukha, H. and A. Benyettou, "A novel framework for defining and submitting workflows to service-oriented systems." *Journal of Information Processing Systems.*, vol. 10, pp. 365–383, 2014.

[4] N. R. BP, "Basics for the construction of phylogenetic trees," *Webmed Central BIOLOGY*, p. 12, 2011.

[5] J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, and G. Dedene, "Formal concept analysis in knowledge processing: A survey on applications," *Expert Systems with Applications*, vol. 40, no. 16, pp. 6538 – 6560, 2013.

[6] R. Abdullah, H. Ibrahim, R. Atan, S. Napis, M. H. Selamat, N. Haslina, and S. Hernazura, "The development of bioinformatics knowledge management system with collaborative environment," *IJCSNS International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp. 309–319, 2008.

[7] E. Antezana, M. Kuiper, and V. Mironov, "Biological knowledge management: the emerging role of the semantic web technologies," *Briefings in Bioinformatics*, 2009.

[8] A. Fiannaca, M. La Rosa, A. Urso, R. Rizzo, and S. Gaglio, "A knowledge-based decision support system in bioinformatics: an application to protein complex extraction." *BMC bioinformatics*, vol. 14, 2013.

[9] R. Maruta, "The creation and management of organizational knowledge," *Knowledge-Based Systems*, vol. 67, pp. 26 –34, 2014.

[10] H. Chen, T. Yu, and J. Y. Chen, "Semantic web meets integrative biology: a survey," *Briefings in Bioinformatics*, vol. 14, no. 1, pp. 109–125, 2013.

[11] K. Sutherland, K. McLeod, G. Ferguson, and A. Burger, "Knowledge driven enhancements for task composition in bioinformatics." *BMC bioinformatics*, vol. 10, 2009.

[12] A. Splendiani, M. Donato, and S. Drghici, "Ontologies for bioinformatics," K. Nikola, Ed. Springer Berlin Heidelberg, 2014, pp. 441–461.

[13] E. Pignotti, P. Edwards, A. Preece, N. Gotts, and G. Polhill, "Enhancing workflow with a semantic description of scientific intent," in *The Semantic Web: Research and Applications*, ser. Lecture Notes in Computer Science, S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds. Springer Berlin Heidelberg, 2008, vol. 5021, pp. 644–658.

[14] Y. Gil, "From data to knowledge to discoveries: Artificial intelligence and scientific workflows," vol. 17, pp. 231– 246, 2009.

[15] M. Aranguren, J. Fernndez Breis, C. Mungall, E. Antezana, A. Gonzlez, and M. Wilkinson, "Oppl galaxy, a galaxy tool for enhancing ontology exploitation as part of bioinformatics workflows," *Journal of Biomedical Semantics*, vol. 4, 2013.

[16] E. Deelman, G. Mehta, G. Singh, M.-H. Su, and K. Vahi, "Pegasus: Mapping large-scale workflows to distributed resources," in *Workflows for e-Science*, I. Taylor, E. Deelman, D. Gannon, and M. Shields, Eds. Springer London, 2007, pp. 376–394.

[17] A. Belloum, R. Cushing, S. Koulouzis, V. Korkhov, D. Vasunin, V. Guevara-Masis, Z. Zhao, and M. Bubak, "Support for cooperative experiments in e-science: From scientific workflows to knowledge sharing," in *Identification of Ligand Binding Site and Protein-Protein Interaction Area*, ser. Focus on Structural Biology, I. Roterman-Konieczna, Ed. Springer Netherlands, 2013, vol. 8, pp. 135–159.

[18] V. Chaudhri, P. Clark, A. Overholtzer, and A. Spaulding, "Question generation from a knowledge base," in *Knowledge Engineering and Knowledge Management*, ser. Lecture Notes in Computer Science, K. Janowicz, S. Schlobach, P. Lambrix, and E. Hyvnen, Eds. Springer International Publishing, 2014, vol. 8876.

[19] J. Girish, G. Arun, and R. Gintaras, "A novel knowledge management system based on workflows," ser. 24th European Symposium on Computer Aided Process Engineering. Elsevier, 2014, vol. 33, pp. 853–858.

[20] Y. Gil, V. Ratnakar, J. Kim, J. Moody, E. Deelman, P. Gonzalez-Calero, and P. Groth, "Wings: Intelligent workflow-based design of computational experiments," *Intelligent Systems, IEEE*, vol. 26, no. 1, pp. 62–72, 2011.

[21] L. Zhang, Y. Wang, P. Xuan, A. Duvall, J. Lowe, Y. Wang, A. Subramanian, P. Srimani, F. Luo, and Y. Duan, "Sesame: A new bioinformatics semantic workflow design system," in *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*, 2013, pp. 504–508.

[22] C. Goble, K. Wolstencroft, A. Goderis, D. Hull, J. Zhao, P. Alper, P. Lord, C. Wroe, K. Belhajjame, D. Turi, R. Stevens, T. Oinn, and D. De Roure, "Knowledge discovery for biology with taverna," in *Semantic Web*, C. Baker and K.-H. Cheung, Eds. Springer US, 2007, pp. 355–395.

[23] J. T. O. Rios, Javier; Karlsson, "Magallanes: a web services discovery and automatic workflow composition tool," *BMC Bioinformatics*, vol. 10, p. 334, 2009.

[24] M. DiBernardo, R. Pottinger, and M. Wilkinson, "Semi-automatic web service composition for the life sciences using the biomoby semantic web framework," *Journal of Biomedical Informatics*, vol. 41, pp. 837–847, 2014.

[25] D. Barseghian, I. Altintas, M. B. Jones, D. Crawl, N. Potter, J. Gallagher, P. Cornillon, M. Schildhauer, E. T. Borer, E. W. Seabloom, and P. R. Hosseini, "Workflows and extensions to the kepler scientific workflow system to support environmental sensor data access and analysis," *Ecological Informatics*, vol. 5, no. 1, pp. 42 – 50, 2010.

[26] M. Sonntag, K. G?rlach, D. Karastoyanova, F. Leymann, P. Malets, and D. Schumm, "Views on scientific workflows," in *Perspectives in Business Informatics Research*, ser. Lecture Notes in Business Information Processing, J. Grabis and M. Kirikova, Eds. Springer Berlin Heidelberg, 2011, vol. 90, pp. 321–335.

[27] E. Mikroyannidi, M. Quesada-Martnez, D. Tsarkov, J. Fernndez Breis, R. Stevens, and I. Palmisano, "A quality assurance workflow for ontologies based on semantic regularities," in *Knowledge Engineering and Knowledge Management*, ser. Lecture Notes in Computer Science, K. Janowicz, S. Schlobach, P. Lambrix, and E. Hyvnen, Eds. Springer International Publishing, 2014, vol. 8876, pp. 288–303.