

# Extending Workflow Management for Knowledge Discovery in Clinico-Genomic Data

Stefan RÜPING<sup>a</sup>, Stelios SFAKIANAKIS<sup>b</sup> and Manolis TSIKNAKIS<sup>b</sup>

<sup>a</sup> *Knowledge Discovery Department, Fraunhofer Institute for Intelligent Analysis and Information Systems, Sankt Augustin, Germany*

<sup>b</sup> *Biomedical Informatics Laboratory, Institute of Computer Science, FORTH, Greece*

**Abstract.** Recent advances in research methods and technologies have resulted in an explosion of information and knowledge about cancers and their treatment. Knowledge Discovery (KD) is a key technique for dealing with this massive amount of data and the challenges of managing the steadily growing amount of available knowledge. In this paper, we present the ACGT integrated project, which is to contribute to the resolution of these problems by developing semantic grid services in support of multi-centric, post-genomic clinical trials. In particular, we describe the challenges of KD in clinico-genomic data in a collaborative Grid framework, and present our approach to overcome these difficulties by improving workflow management, construction and managing workflow results and provenance information. Our approach combines several techniques into a framework that is suitable to address the problems of interactivity and multiple dependencies between workflows, services, and data.

**Keywords.** Biomedical Semantic Grid, Grid Workflows, Knowledge Discovery

## Introduction

Recent advances in research methods and technologies have resulted in an explosion of information and knowledge about cancers and their treatment. The ability to characterize and understand cancer is growing exponentially based on information from genetic and protein studies, clinical trials, and other research endeavours. The breadth and depth of information already available in the research community at large, present an enormous opportunity for improving our ability to reduce mortality from cancer, improve therapies and meet the demanding individualization of care needs [1,2]. Because our knowledge of this domain is still largely rudimentary, investigations are now moving from being *hypothesis-driven* to being "data-driven" with analysis based on a search for biologically relevant patterns.

While these opportunities exist, the lack of a common infrastructure has prevented clinical research institutions from being able to mine and analyze disparate data sources. Moreover, clinical researchers or molecular biologists often find it hard to exploit each other's expertise due to the absence of a cooperative environment which enables the sharing of data, resources or tools for comparing results and experiments, and a uniform plat-

form supporting the seamless integration and analysis of disease-related data at all levels [2]. Similarly, the lack of a unifying architecture is proving to be a major roadblock to a researcher's ability to mine different databases. The vision of the Advancing Clinico-Genomic Trials on Cancer (ACGT) integrated project is to contribute to the resolution of these problems by developing a semantic grid services in support of multi-centric, post-genomic clinical trials, and thus enabling discoveries in the laboratory to be quickly transferred to the clinical management and treatment of patients. In achieving such an objective, extensive use of ontologies and metadata are required. This will be provided by developing a master ontology on cancer and implementing suitable metadata registries, which provide semantically rich information about available data and computational services.

In this paper, we present our approach on how to improve the support for KD in a grid environment, in particular with respect to the requirements put up by the domain of post genomic clinical trials. While we are focusing on the data miner's view on data analysis, we always keep in mind the cross-disciplinarity of the research field (indeed, we will see that a thorough domain understanding and interaction with the end-user is a crucial aspect of data mining). The remainder of the paper is organized as follows: Section 1 gives an overview over the field of KD and lists several requirements that are not satisfactorily supported by current Grid environments. Section 2 gives an overview of the ACGT architecture and an assortment of techniques to improve the support for KD within existing Grid architecture are presented. Section 3 gives an overview of the related work and Section 4 concludes.

## **1. Knowledge Discovery**

Knowledge Discovery has been described as "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [12]. A model for a typical data mining process has been proposed by the CRISP project [11]. An interactive, iterative nature of the process is generally characterized as a main property of data mining, see e.g. [13]. The reason is that data mining has an experimental nature, meaning that often several approaches are tried out in parallel before the best one is selected.

### *1.1. Knowledge Discovery in Clinico-Genomic Data*

Data mining methodology and technology has been developed to date for classical business, finance, and customer-oriented application domains. Such domains are characterized by the availability of large quantities of data in an attribute-value based representation, high ratio of examples over attributes in the data sets, and weak background knowledge about the underlying entities and processes.

For biomedical data these conditions do not hold. With high-throughput technologies like microarrays or mass-spectrometry, masses of genomic/proteomic data are produced – about  $10^4$  gene, or mass/charge (m/Z) features per patient. In addition, legal, ethical and practical limitations make it cumbersome to acquire a high number of patients in a clinical trial – a typical (preclinical) cohort may contain only about 100-200 patients. Under these conditions, standard statistical methods are likely to over-fit the structures

in the data, such that a high amount of domain knowledge is needed to guide the analysis and guarantee the validity of the extracted knowledge [3].

It follows that the challenges of KD in bio-medical data differs significantly from the original problems of data analysis that prompted the development of Grid technologies: instead of the analysis of huge data sets, the problem here lies in the analysis of many small data sets with a plethora of possible analysis workflows. The central factor here is to make effective use of the distributed knowledge of the involved research communities in order to compensate the low statistical significance which results from small sample sizes. Valuable kinds of knowledge include:

- *Knowledge about the semantics of the data*: it is well known that in data mining finding an optimal representation of the data is central for obtaining good results. That is, great care must be taken in the step of pre-processing, e.g. feature selection and construction [8].
- *Knowledge about the plausibility of results*: when there is not enough statistical information about the validity of a hypothesis, one can look for external evidence for or against this hypothesis in the scientific literature, where usually much more knowledge is available than what is encoded in the specific data set. To make use of this knowledge, the interpretability of the models must be ensured [14], and text mining technologies [16] must be incorporated.
- *Knowledge about analysis workflows*, in particular about which workflow is optimal for a given problem. This problem is an instance of the field of workflow mining [9] and related approaches [10].

Therefore, the main challenge for KD in clinico-genomic data is the sharing of knowledge, either in the form of the integration of existing knowledge to design and select appropriate analysis tools, or to manage the discovered knowledge in order to make it available to the research communities. In particular, the following five requirements for a workflow-based environment have to be met:

- *Data is central*: workflow execution gives a central position to the services that process the data and not to the data itself. The reason is that usually these functions remain constant, while the data they are executed on varies from execution to execution. In KD, however, the fixed point is the data set which is to be analysed. Several functions can be executed on the data set to solve the research question, and it is not a-priori clear which approach (workflow) will be optimal. The analyst will typically execute a large variety of functions on the same data set, and readily discard any approach that fails to answer his question. Hence, data and results should play a much more prominent role in workflow construction for KD than in usual approaches.
- *The need to support interactivity*: as stated before, KD is inherently an interactive process. Workflow enactment must support this nature of the process in order to match the trial-and-error way of working in data analysis. For example, a typical step in KD is to check the intermediary results of a workflow and then decide whether to continue the workflow or to modify it and re-start the analysis.
- *Managing dependencies between discovered knowledge and experiments*: as a result from the data- and hypothesis-driven working style, several data sets, workflows, workflow executions, and results will be generated during a single KD process. Traditionally, it is the responsibility of the analyst to keep track of all

these instances and their relationships, e.g. which workflow generated which result, what methods have been tried out on a data set, etc. With the possibilities of collaborative and massively parallel data analysis on the grid, keeping track of all these dependencies becomes infeasible if it is not automatically supported by the system.

- *Mining-specific provenance information and meta data needs to be supported.* Important information that relates to data mining are for example the mapping of services to stages in the CRISP process, a description of data in terms of its role in the mining process (e.g. input data, transformed data, model), the relation between different results (A is better than B), and information about the applicability of services to relevant data types (e.g. nominal, or numeric values, text data, images).
- *Support of building blocks:* it can often be found that a certain sequence of operators that is part of a workflow is meaningful on its own, for example because it corresponds to a phase in the CRISP model, or because a certain operator cannot be applied without a fixed sequence of pre-processing steps. In these cases, a method for the composition of a sequence of services into a single service that can be appropriately marked up and retrieved by a semantic discovery method is very useful.

In the following chapter, we will propose approaches to meet these requirements based on existing Grid architectures.

## 2. Architecture

The current state of the art architectural style for building distributed applications is the so called Service Oriented Architecture (SOA) where the building block is the notion of Services. A service is usually defined as “a software system designed to support interoperable machine-to-machine interaction over a network”<sup>1</sup> typically using SOAP as the communication protocol for exchanging XML messages complying to an agreed contract expressed in WSDL. The SOA methodology has been also adopted by the Grid computing community and we currently see a convergence of Web Services and Grid Services in the context of Open Grid Service Architecture [27]. An architecture based on services and their interactions has therefore been defined for the development of the ACGT infrastructure. An overview of this layered architecture is depicted in Figure 1.

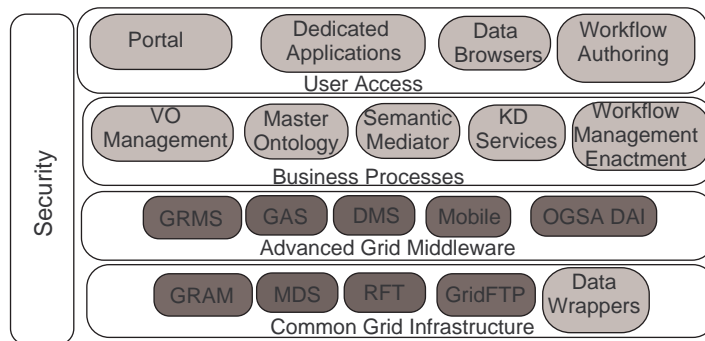
In the general ACGT environment the workflow authoring and management tasks play a central role to support the KD process. Therefore, a more detailed architecture has been defined to support the publication, discovery, invocation, and management of workflows and it will be further elaborated during the course of the project. Based on the discussion in Section 1 we have come up with the following approaches:

### 2.1. Workflows as services

The Workflow Management Coalition (WFMC) defines a workflow as “The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural

---

<sup>1</sup><http://www.w3.org/TR/ws-arch/>



**Figure 1.** An overview of the ACGT architecture

rules". In other words a workflow consists of all the steps and the orchestration of a set of activities that should be executed in order to deliver an output or achieve a larger or more sophisticated goal. In essence a workflow can be abstracted as a composite service, i.e. a service that is composed by other services that are orchestrated in order to perform some higher level functionality. This idea will be further extended and become concrete during the ACGT project.

On the practical side we are planning to support the implementation of a service that accepts a workflow description in some workflow description language and automatically generates a WSDL that describes this workflow. This will make the workflows easily accessible and reused from the rest of the ACGT environment using familiar web services technologies.

## 2.2. *Ontologies and Semantic Descriptions*

The importance of annotation of services and workflows with semantic descriptions cannot be overemphasized. An adequate and convenient way of annotating workflows with metadata is of paramount importance for their posterior discovery among repositories of useful workflows based on their purpose, internals and field of application characteristics.

In trying to respond to this need in ACGT we are developing a domain ontology for cancer research that will enable the semantic integration of data and services. This ACGT Master Ontology will serve as a common reference for concepts related to cancer and their relationships that is both, human-understandable and machine-understandable. In addition to the Master Ontology another ontology specific for the description of the services details is needed. This Service Ontology is necessary for the description of the data formats, the service functionalities, provenance information, and so on. These issues have already been tackled in the myGrid project [30] and we will base our implementation on their work.

## 2.3. *Data as services*

In a scientific workflow data is the information that flows between services and triggers their execution and also the final outcome of a workflow execution. In addition to workflows being services we also promote the modeling of data, be it intermediate or final

results of a workflow, as services. We consider this approach to be feasible, since each data source that is somehow accessible can be considered as a service that when invoked returns the data that it encapsulates. This is a common phenomenon in the World Wide Web where accessing a HTTP URI does not usually disclose if the retrieved content is dynamically generated by some web service or not.

It is also desirable, since the handling of data as basic services will enable a unified approach for the integration of data and services into workflows. This means that we can have the same infrastructure for annotating, searching, matching, and composing data, services, and workflows. It also means that data are "first class citizens" in the ACGT environment and not just the channels of communication between the services.

#### 2.4. Breakpoints

In order to improve the interactivity of Grid workflows, we propose the implementation of a so-called Breakpoint Service. The task of this service when it is executed is to store the state of the workflow execution (e.g. the intermediate results, the execution's control point) as a continuation [39] to a user-specific data repository. It will then construct a new *continuation workflow* that starts with extracting this input data from the data repository and continue with the workflow that the Breakpoint Service is a part of. This new workflow will be entered into the workflow repository, with appropriate information linking it to the workflow data and the original workflow it was derived from, and converted into a service to allow an easy access. Finally, the Breakpoint Service will suspend the running workflow.

When the user is notified that his initial workflow has stopped (which will be done by the usual Grid management services), he can then look for the new workflow-representing service (using the usual semantic service discovery and notification services) and decide to inspect the intermediate results of the workflow from his data repository, continue the workflow, i.e. execute the newly constructed continuation workflow, load the continuation workflow into the workflow editor to make appropriate changes, or load the original workflow into the workflow editor to make appropriate changes and restart.

In this way, a significant extension of the interactivity and flexibility of workflows can be achieved within existing architectures.

#### 2.5. Hyperlinked Provenance Information

Provenance information is the key to organizing information in a collaborative, Grid-based system. This becomes an even greater necessity when improved interactivity can result in changing workflows on the fly and cloning running workflows to explore alternatives in parallel. A hyperlinked presentation of information has been proposed as a tool for better supporting the collaboration in scientific communities [15]. In combination with the idea of representing data as services and workflows as services, provenance information can be viewed as a graph of services, with edges representing several types of provenance. For example, edges can represent the relations "produced-by", "part-of", "derived-from", "input-of" etc.

The user can navigate in such a graph just like surfing the web, where each service corresponds to a web page, which could be available in different representations to adapt

to the needs of different user groups, and edges correspond to links. Each entity (e.g. service, data) is identified by a HTTP URI to provide identification, retrieval, and linking facilities for constructing a web of data and metadata in accordance with the Semantic Web vision [37].

The employment of this prototypical web of provenance information on the ubiquitous infrastructure of World Wide Web will have the advantage of user familiarity while offering a basis for more advanced features such as a bridge to the Semantic Web through GRDDL [34] transformations for example.

### 3. Related Work

A practical approach to provide standard data mining services across a Grid infrastructure is the integration of standard statistical and data mining toolkits, such as Weka [4] or R [5]. This approach was followed by the DataMiningGrid project<sup>2</sup>, which focused on facilitating large-scale, distributed data mining operations, but less to support interconnected data mining tasks as they are prominent in the analysis of biomedical data. Hence, this project needs to be complemented with user-oriented services to structure and manage data mining results.

The Smart Tea project [18] and its successor myTea [19], have introduced the concept of an electronic lab book. The idea is to bridge the gap between workflow execution on the grid and work done in the lab in order to enable meaningful interrogation of the process and controlled sharing of the results. Other authors have proposed approaches for linking and browsing lab information as well, most notable the approach for provenance information in the myGrid project [21,20].

IDEA [17] is an Intelligent Discovery Assistant, which provide users with systematic enumerations of valid KD processes, and effective rankings of these valid processes by different criteria, to facilitate the choice of KD processes to execute. It is based on an ontology of KD operators, which guides the combination of services into valid workflows. In contrast to the ACGT approach, IDEA does not make use of a refined ontology on the data, because it is meant to be a general-purpose tool and it cannot be assumed that such an ontology exists.

The Mining Mart system [10] focusses on setting up and re-using best-practice cases of preprocessing data stored in very large databases. A meta-data model named M4 is used to declaratively define and document both, all steps of such a pre-processing chain and all the data involved. The idea of Mining Mart is to publish best-practice cases on the internet. However, the system only contains information about the workflows, not about models, because the system was developed in a business context, where models are usually private information

There is a lot of research in the area of workflows and service composition. There are plenty of available tools for authoring scientific workflows in the Grid [28,26,29] with different characteristics supported by each of them [35].

In myGrid's Taverna Workbench<sup>3</sup> services can be annotated with semantic descriptions based on ontologies and later discovered based on these descriptions. The myGRID ontology [30] will be used also in our work but in linkage with our specific domain ontol-

---

<sup>2</sup><http://www.datamininggrid.org/>

<sup>3</sup><http://sourceforge.net/projects/taverna>

ogy for cancer. The information about the workflows, their enactment, and provenance is kept in the myGrid Information Repository (mIR) which is a UDDI compliant registry. The provenance logs are also annotated with concepts from the domain ontology and a web of provenance documents, services, and workflows can be built [21]. The sharing and reuse of workflows in myGrid has recently been enabled by the design of myExperiment<sup>4</sup>. In our case the hyperlinked provenance information is based on the semantic web approach where each identified entity is given a URI that can be “dereferenced” and easily retrieved by both humans and software. Nevertheless, the general concepts are the same and we are investigating methods for reusing parts of the myGrid infrastructure .

Continuations have been proposed in the past as a means for maintaining conversational state in web interactions [38]. The applicability of continuations in workflow enactment has also been described in the past [40] but to our knowledge there are few workflow tools to support them (e.g. BpmScript<sup>5</sup>). An interaction capability has been recently added to the Taverna workbench through the Interaction Service. An interaction service when invoked sends an e-mail to the user with the information needed to stop or continue the workflow by the means of HTTP links.

For the semantic annotation of services we consider OWL-S<sup>6</sup> to be highly relevant, since it can be used to describe workflows as “composite processes” through its process model. OWL-WS (“OWL for Workflows and Services”) [36] is a workflow and service ontology supported by the NextGRID project that is based on OWL-S with various extensions to provide advanced features like the abstract definition of workflows so that a task’s bindings (endpoints, etc.) can be specified/located at run time, and higher order workflows so that a task can have a workflow as input and return another workflow as output. Finally the semantic annotation of WSDL [41] is a way to put inline references to user specified ontologies in the WSDL documents. Although the combination of data and metadata in a single entity has clear advantages, we don’t consider the current breed of tools to be mature enough to support this “intrusive” approach.

#### 4. Conclusions

Knowledge discovery in clinico-genomic data exhibits interesting challenges for workflow management and execution, which are not fully supported by current Grid workbenches. We have presented an overview of the roadmap in the ACGT project with regards to the KD process in clinico-genomic data using workflows. We propose the combination of several Grid techniques in order to provide an interactive, easy-to-use, yet expressive environment that supports the needs of both domain scientists, e.g. clinicians and biologists, and data miners. These techniques will be further extended and realized in the course of the project taking into account the prior art in the relevant fields of workflow management and enactment and the semantic composition of services while following the user requirements elicitation process.

---

<sup>4</sup><http://www.myexperiment.org/>

<sup>5</sup><http://www.bpmscript.org/>

<sup>6</sup><http://www.daml.org/services/owl-s/>



## Acknowledgements

The authors wish to thank to the ACGT consortium for their contributions with various ideas on which the ACGT project was developed. The ACGT project is funded by the European Commission (Contract No. FP6/2004/IST-026996).

## References

- [1] K. Yurkewicz, *Accelerating Cancer Research*, Science Grid, June 21, 2006. Available at [http://www.interactions.org/sgtw/2006/0621/cabig\\_more.html](http://www.interactions.org/sgtw/2006/0621/cabig_more.html).
- [2] K. H. Buetow, *Cyberinfrastructure: Empowering a "Third Way" in Biomedical Research*, Science Magazine, vol. 308, no. 5723, pp. 821-824, 2005.
- [3] M. May, G. Potamias, S. Rüping, *Grid-based Knowledge Discovery in Clinico-Genomic Data*, Proc. Int. Symposium on Biological and Medical Data Analysis (ISBMDA), 2006.
- [4] I. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, 2000.
- [5] R Development Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2005, ISBN 3-900051-07-0.
- [6] PMML Schema Version 2.1: <http://www.dmg.org/v2-1/GeneralStructure.html>
- [7] OWL-S; <http://www.daml.org/services/owl-s/>
- [8] H. Liu and H. Motoda (eds.), "Feature Extraction, Construction and Selection: A Data Mining Perspective", Kluwer, 1998.
- [9] W. van der Aalst et al., *Workflow Mining: a Survey of Issues and Approaches*, in: Data and Knowledge Engineering, 47 (2), pp. 237-267, 2003.
- [10] K. Morik and M. Scholz, *The MiningMart Approach to Knowledge Discovery in Databases*. in: Ning Zhong and Jiming Liu (eds.), Intelligent Technologies for Information Analysis, Springer, pp. 47 - 65, 2004.
- [11] C. Chapman et al., *CRISP-DM 1.0*, <http://www.crisp-dm.org>, 2000.
- [12] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, *From Data Mining to Knowledge Discovery: An Overview*. In Advances in Knowledge Discovery and Data Mining, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Menlo Park, Calif.: AAAI Press, 1-30, 1996.
- [13] R. Brachman, T. Anand, *The Process of Knowledge Discovery in Databases: A Human-Centered Approach*. In Advances in Knowledge Discovery and Data Mining, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Menlo Park, Calif.: AAAI Press, pp. 37-85, 1996.
- [14] S. Rüping, *Learning Interpretable Models*, Ph.D. dissertation, University of Dortmund, Germany, 2006.
- [15] C. Goble, O. Gocho, P. Alper, D. De Roure, *e-Science and the Semantic Web: A Symbiotic Relationship*, 9th Int. Conf. on Discovery Science (DS2006), Springer, pp. 1-12, 2006.
- [16] R. Feldman, J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2006.
- [17] A. Bernstein, F. Provost, *An Intelligent Assistant for the Knowledge Discovery Process* Proc. IJCAI 2001 Workshop on Wrappers for Performance Enhancement in KDD, 2001.
- [18] M. Schraefel, G. Hughes, H. Mills, G. Smith, T. Payne, J. Frey, *Breaking the Book: Translating the Chemistry Lab Book into a Pervasive Computing Lab Environment*, in Proceedings of CHI 2004, Vienna, Austria, 2004.
- [19] A. Gibson, R. Stevens, R. Cooke, S. Brostoff, M. Schraefel, *myTea: Connecting the Web to Digital Science on the Desktop*, Technical Report.
- [20] J. Zhao, C. Goble, R. Stevens, S. Bechhofer, *Semantically Linking and Browsing Provenance Logs for e-Science*, in International Conference on Semantics of a Networked World (IC-SNW ), pp. 157-174, Springer, 2004
- [21] J. Zhao, C. Goble, M. Greenwood, C. Wroe, R. Stevens, *Annotating, linking and browsing provenance logs for e-Science*, in 1st Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, Sanibel Island, Florida, USA, 2003.
- [22] Ajax: A New Approach to Web Applications. Adaptive Path (2005-02-18). <http://www.adaptivepath.com/publications/essays/archives/000385.php>

- [23] Business Process Execution Language for Web Services version 1.1 <http://www-128.ibm.com/developerworks/library/specification/ws-bpel/>
- [24] ECMAScript Language Specification, Standard ECMA-262, 3rd edition (December 1999) <http://www.ecma-international.org/publications/files/ecma-st/ECMA-262.pdf>
- [25] I. Foster, C. Kesselman, S. Tuecke. *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*. International J. Supercomputer Applications, 15(3), 2001.
- [26] I. Altintas, C. Berkley, E. Jaeger, M. Jones, B. Ludäscher, S. Mock *Kepler: An Extensible System for Design and Execution of Scientific Workflows*, 16th Intl. Conf. on Scientific and Statistical Database Management (SSDBM'04), 21-23 June 2004, Greece.
- [27] J. Nick, S. Tuecke, *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. Open Grid Service Infrastructure WG, Global Grid Forum, 2002.
- [28] Tom Oinn, Mark Greenwood, Matthew Addis, M. Nedim Alpdemir, Justin Ferris, Kevin Glover, Carole Goble, Antoon Goderis, Duncan Hull, Darren Marvin, Peter Li, Phillip Lord, Matthew R. Pocock, Martin Senger, Robert Stevens, Anil Wipat and Chris Wroe. *Taverna: Lessons in creating a workflow environment for the life sciences* accepted for publication in Concurrency and Computation: Practice and Experience Grid Workflow Special Issue
- [29] S. Majithia, M. S. Shields, I. J. Taylor, and I. Wang. *Triana: A Graphical Web Service Composition and Execution Toolkit*. Proceedings of the IEEE International Conference on Web Services (ICWS'04), pages 514-524. IEEE Computer Society, 2004.
- [30] C Wroe, R Stevens, CA Goble, A Roberts, and M Greenwood. *A suite of DAML+OIL ontologies to describe bioinformatics web services and data*. In International Journal of Cooperative Information Systems, 12(2):197-224, March 2003.
- [31] SPARQL Query Language for RDF, W3C Working Draft 4 October 2006, <http://www.w3.org/TR/rdf-sparql-query/>
- [32] SPARQL Protocol for RDF, W3C Candidate Recommendation 6 April 2006, <http://www.w3.org/TR/rdf-sparql-protocol/>
- [33] Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation 10 February 2004 <http://www.w3.org/TR/rdf-concepts/>
- [34] Gleaning Resource Descriptions from Dialects of Languages, W3C Working Draft 24 October 2006, <http://www.w3.org/TR/grddl/>
- [35] Jia Yu, Rajkumar Buyya, *A Taxonomy of Workflow Management Systems for Grid Computing*, Technical Report, GRIDS-TR-2005-1, Grid Computing and Distributed Systems Laboratory, University of Melbourne, Australia, March 10, 2005.
- [36] Stefano Beco, Barbara Cantalupo, Ludovico Giammarino, Mike Surrige and Nikolaos Matskanis *OWL-WS: A Workflow Ontology for Dynamic Grid Service Composition*, accepted to the 1st IEEE International Conference on e-Science and Grid Computing, Dec. 5 - 8, 2005, Melbourne, Australia
- [37] Tim Berners-Lee *Linked Data* Design Issue, <http://www.w3.org/DesignIssues/LinkedData.html>
- [38] Christian Queinnec. *Inverting back the inversion of control or, continuations versus page-centric programming*. Technical Report 7, LIP6, May 2001.
- [39] R. Kelsey, W. Clinger, J. Rees (eds.), *Revised5 Report on the Algorithmic Language Scheme*, Higher-Order and Symbolic Computation, Vol. 11, No. 1, August, 1998
- [40] D. Manolescu. *Workflow Enactment with Continuation and Future Objects* OOPLSA'02, Seattle, WA, 2002. <http://citeseer.ist.psu.edu/manolescu02workflow.html>
- [41] Semantic Annotations for WSDL W3C Working Draft 28 September 2006 <http://www.w3.org/TR/sawsdl/>
- [42] Workflow Management Coalition Workflow Reference Model, <http://wfmc.org/standards/model.htm>