Front-ends to Biomedical Data Analysis on Grids

Shayan SHAHAND [a] Mark SANTCROOS [a] Yassene MOHAMMED [a,d]
Vladimir KORKHOV [a,c] Angela C. M. LUYF [a] Antoine H. C. VAN KAMPEN [a,b]
Silvia D. OLABARRIAGA [a] [1]

[a] *Bioinformatics Laboratory, Department of Clinical Epidemiology Biostatistics and Bioinformatics, Academic Medical Center, PO Box 22700, 1100 DE Amsterdam, The Netherlands*
[b] *Biosystems Data Analysis, Swammerdam Institute for Life Science, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
[c] *Faculty of Applied Math and Control Processes, St.Petersburg State University, Russia*
[d] *Regional Computing Center for Lower Saxony and L3S, University of Hannover, Germany*

Abstract.

The e-infrastructure for bioscience (e-BioInfra) is a platform integrating various services and middleware to facilitate access to grid resources for biomedical researchers at the Academic Medical Center of the University of Amsterdam. In the past six years the user interfaces with the e-BioInfra have evolved from command-line interfaces to a Java desktop application, and later to an easy-to-use web application for selected biomedical data analysis. This evolution represents improvements to accommodate the requirements of a broader range of biomedical researchers and applications. In this paper we present the current user interfaces and analyse their usage considering the typical biomedical data analysis on the e-BioInfra, the roles assumed by the users in the various phases of data analysis life-cycle, and the user profiles. We observe that in order to support a wide spectrum of user profiles, with different expertise and requirements, a platform must offer a variety of user interfaces addressed to each user profile.

Keywords. science gateway, grid computing, biomedical research, e-science, grid user interface

Introduction

Biomedical research applications become increasingly compute-intensive and data-intensive, and as such can benefit from distributed computing infrastructures (DCI), e.g., grid computing. The *e-Bioscience Infrastructure (e-BioInfra)* was introduced by Olabarriaga et al. [20] to support medical image analysis on the Dutch e-science grid infrastructure. Since 2005 it has evolved into a platform integrating various services and

---

middleware to facilitate access to grid resources. Currently it is adopted by biomedical researchers at the Academic Medical Center (AMC) of the University of Amsterdam (UvA), and a similar approach is also adopted at CREATIS, Lyon, France, for medical imaging research [5].

The platform evolution from a "low hanging fruit" up to a "user-ready" phase is presented in [19]. In the initial phase only command-line interfaces (CLI) were available. In the user-ready phase a service-oriented approach was adopted to encapsulate the complexity of accessing grid resources into high-level services that could be accessed via the Virtual Resource Browser (VBrowser) which was the single point of access to these services. Workflows are used as basic technology to perform grid computation, enabling applications to be executed on the grid. Our experience shows that even though this technology facilitates data analysis on grids, adopting it remains difficult for biomedical researchers. They do not want to get involved in the grid computing complexities, neither are interested in knowing all technical details. They just want to use applications developed and gridified by others to analyse their data. For these users, a web interface has been recently added to the e-BioInfra. We realised, however, that all the three types of user interfaces remained in use because users have different preferences, requirements and expertise. In this paper we analyse the usage of the various interfaces and attempt to characterise the user profiles for each type.

The paper is organised as follows: In section 1 we present the e-BioInfra architecture and the functionality of its components. In section 2 we present an analysis of the phases in the life-cycle of biomedical data analysis, roles, usage patterns and user profiles we observed. We finish the paper with a brief overview of related work in section 3, and conclusion and future work in section 4.

1. An e-Infrastructure for Bioscience: e-BioInfra

In this section we present the latest developments of the improved system architecture as an update to the description in [20]. The main differences concern the web applications layer and a dedicated solution to transport data to/from grid resources. These were necessary to accommodate the new requirements of a broader range of biomedical research applications and users. We follow a bottom-up approach to describe the building blocks of the e-BioInfra architecture as illustrated in figure 1. Note that the components marked with star are new to the system architecture described in [20].

1.1. Grid Fabric and Services Layers

The bottom layer is the grid fabric composed of hardware resources, including Compute Elements (CE), Worker Nodes (WN) and Storage Elements (SE). Resources are provided by the Dutch e-Science Grid (BiG Grid project[2]) and currently consist of around 7,000 cores and 60 TB of storage. Access to the resources is granted via membership to the VLEMED Virtual Organization (VO), using Grid Security Infrastructure (GSI) authentication and authorization. Users in possession of a valid grid certificate are allowed to use these resources.
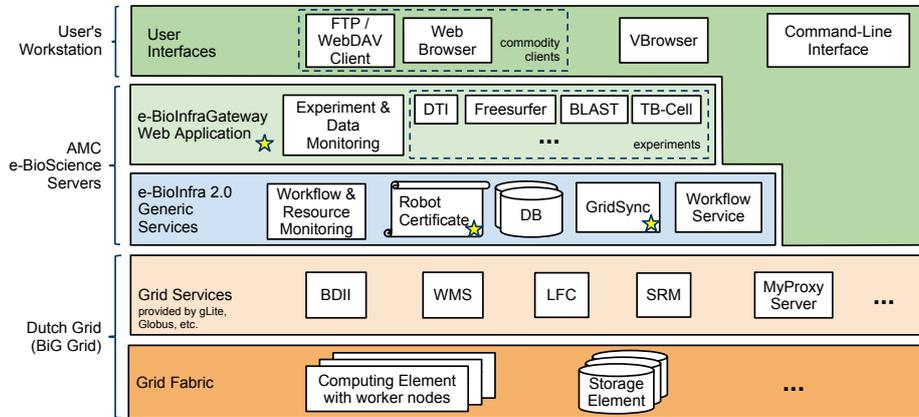
---

[2]http://www.biggrid.nl

Figure 1. e-BioInfra layered architecture: grid fabric and services, generic services, e-BioInfraGateway web application and user interfaces. Components marked with star are new in the e-BioInfra.

The middleware components form the grid services layer, for instance, gLite Workload Management System (WMS), Logical File Catalog (LFC), Storage Resource Management (SRM), and Berkeley Database Information Index (BDII). These services are operated by the BiG Grid project.

1.2. e-BioInfra Generic Services

The generic services layer offers a higher level of abstraction to interact with the grid services, wrapping them to facilitate usage. The e-BioInfra generic services include:

Workflow Service   web-service that wraps the MOTEUR workflow management engine [9], which is used to enact workflows on the grid resources. Workflows are described in the GWENDIA language [15]. As illustrated in figure 2, MOTEUR uses WMS or the DIANE[3] pilot job framework [16] to execute workflow tasks as jobs on the grid. This web-service also provides status information about submitted workflows. The service can be invoked directly through various user interfaces, as discussed in section 1.4.

GridSync   this service is responsible for transporting data and results between the hospital network and the grid storage as illustrated in figure 3, serving novice users who are not familiar grid protocols nor have a grid certificate. GridSync runs on a separate server within the hospital network and it initiates all transfers, for both incoming and outgoing data. All data transported by this service is assumed to be anonymized to satisfy privacy regulations. This service is tightly coupled with the e-BioInfraGateway, meaning that it is configured to transfer data to and from pre-configured locations on the grid storage resources. Note that the data staging is only a stepping stone for the files and it is not to be used for permanent storage; all files on the data staging server are considered volatile and removed regularly with user awareness. The users can choose a familiar data transfer client, e.g., sFTP or WebDAV client, to manage their data on this server, and the service takes care of the synchronisation with the grid storage. The basic principle is that all
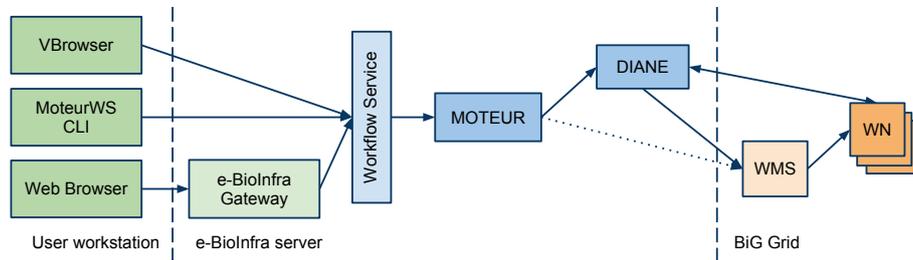
---

[3]http://cern.ch/diane

3

Figure 2. Workflow web-service is used through different user interfaces to submit workflows to the grid using DIANE or the WMS directly (dotted line).

files put by the user in a particular directory are automatically mirrored to a pre-defined directory on the grid resources. Similarly, all files resulting from grid jobs are stored in a given directory and mirrored back to the data staging server automatically.
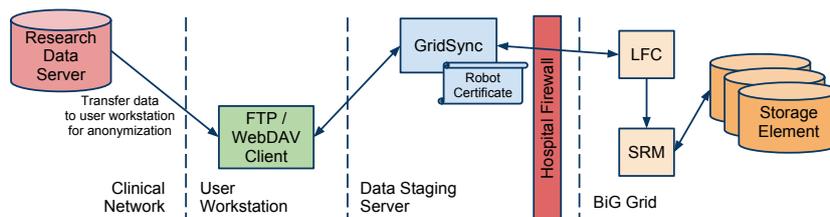

Figure 3. GridSync transfers data to/from the grid

Robot Certificate a valid grid certificate is used to generate a grid proxy which is essential for using grid resources. A robot certificate is used periodically to generate a grid proxy automatically on the MyProxy server. This generated grid proxy is then used by the GridSync service to synchronise directories with the grid storage, or by the e-BioInfraGateway web application to submit workflows to the workflow service. The robot proxy is only accessible by certain e-BioInfra services (i.e., users cannot use it directly), to enable authorized users who do not have a grid certificate use the grid infrastructure.

Workflow and Resource Monitoring component is used to monitor and track user activities, for example providing detailed information about the workflows and jobs, DIANE pilot agents; It also provides information about available storage and resource status for provisioning and support purposes.

Databases store structured information about users, experiments, scientific datasets, workflows, etc. It is used by the workflow and resource monitoring component and the e-BioInfraGateway web application (see section 1.3).

1.3. e-BioInfraGateway Web Application

The e-BioInfraGateway web application[4] is built on top of the e-BioInfra generic services. It is developed using the Spring framework[5], providing a generic substrate into which workflows can be integrated and executed using a web interface. In addition to submitting the integrated workflows, a.k.a. *experiments*, the e-BioInfraGateway also generates a predefined directory structure for input data and results as well as identifiers to facilitate experiment management. Each experiment is implemented as a Spring Web Flow making the e-BioInfraGateway easily extensible.

After successful registration, the user is granted access to the e-BioInfraGateway, using username and password. It is not necessary for the user to be in possession of a grid certificate because all of the interactions between the web application and the grid infrastructure are authenticated and authorized by the robot proxy. The information about the experiment and file owners is stored in the database. In addition, all user activities are logged for accountability purposes.

To start a new experiment, first the user is directed to a secured personal FTP directory where he/she can upload the input data; optionally the user can upload the input data using VBrowser directly to the grid storage. The data is pushed automatically to the grid storage by the GridSync service. When ready, the user can start the experiment via the web interface. After an experiment finishes successfully, the GridSync service fetches the results automatically from the grid storage and stores them in a predefined directory on the server.

Currently there are five experiments integrated into the e-BioInfraGateway for applications in medical imaging and DNA sequencing:

- analysis of magnetic resonance Diffusion Tensor Imaging (DTI) in particular the construction of a "brain atlas" for a particular study (see more details in [4]);
- segmentation of functional Magnetic Resonance Imaging (MRI) data with the Freesurfer toolbox [7];
- DNA sequence alignment with the BLAST toolbox [14];
- analysis of T/B-cell variation in different organisms [11];
- comparing genomes of related species based on the occurrence of common genes.

The last two experiments have been integrated recently by external developers, showing that the gateway can be extended with other experiments in a straightforward manner.

Note that, whereas there is a clear trend to use frameworks to implement portal interfaces (see section 3), such technology was not our first choice due to the steep learning curve. Instead we chose for a light-weight solution to implement the web application.

1.4. User Interfaces

Users may choose between three types of interfaces to interact with the platform based on their roles, expertise and requirements:

---

[4]http://orange.ebioscience.amc.nl/ebioinfragateway
[5]http://www.springsource.org

Command-Line Interfaces (CLI) are the most flexible interfaces to the platform, but they require technical knowledge and experience. They are mostly used by advanced users who like to take control of all of the parameters in the system, those who wish to use them in scripts, or for debugging purposes. Examples are the MOTEUR web-service client (MoteurWS CLI) and the gLite [13] CLI clients. CLIs require the proper software stack to be installed on the system (e.g., gLite), so they are typically executed in servers with the gLite user interface (UI) installation. This also requires the user to possess a valid grid certificate.

Virtual Resource Browser or VBrowser[6], is a Java desktop application to access and manage local and remote (grid) resources [18]. It is equipped with a number of plug-ins, e.g., MOTEUR plug-in to submit workflows to the workflow service. This interface is typically used to manage data on the grid, start workflows and view some type of results (e.g., images). The VBrowser is installed on the user's workstation and configured to access the remote resources, and it requires the user to possess a grid certificate to access grid resources. Because of the GUI, the VBrowser is easier to use, however its configuration and operation still requires some knowledge of grid usage and jargon.

Commodity clients are general purpose web browsers, FTP and WebDAV programs can be used to access the e-BioInfraGateway and GridSync services respectively. These interfaces are publicly available and usually pre-installed on every workstation. The user does not need a grid certificate to use these interfaces, however authentication through username and password is essential. These user interfaces are the easiest for biomedical researchers without grid experience because they shield the grid completely from the user's perspective. However, they offer limited functionality, for example, only pre-defined workflows can be started from a web interface.

2. Usage of the e-BioInfra

The e-BioInfra is being adopted by a growing community of biomedical researchers from medical imaging and genomics to perform data analysis, as well as members of the development and support team. These users have heterogeneous interests and backgrounds, and also display different usage patterns. In this section we present an analysis of the usage following closely the structure suggested in [21], which identified phases of the life-cycle of a typical medical image analysis and a set of roles and tasks for each phase. Here we revisit that work in the broader context of data analysis in biomedical research using the e-BioInfra, reformulating the phases in the life-cycle and the different roles. Finally we analyse the user profiles taking into account the life-cycle, roles and tasks performed by the users.

2.1. Life-cycle Phases

First the components of the workflows are developed, either in-house or by third-party collaborators, and then the data analysis workflow is developed; this is called the *development phase*. The parameters of the workflow and its components are then optimized

---

[6]http://www.vl-e.nl/vbrowser

in the *optimization phase* and evaluated to certify that it is working accurately for representative data-sets in the *evaluation phase*.

When workflows are ready, they can be shared with a larger user community. To achieve this, the workflow can be published in some repository (directory on some shared storage such as the LFC) or embedded into some customized user interface (integrated as experiment in the e-BioInfraGateway). We coin this as *deployment phase*; from this point on the workflow enters the *production phase* (clinical routine phase in [21]) in which it is used routinely for data analysis by users that are not necessarily aware of the technical details.

2.2. Roles and Tasks

In our analysis we could roughly identify the same roles and tasks identified in [21]. The component developer builds new data analysis methods that can be combined into workflows by the workflow developer. We observed that the same person also optimizes the parameters of the workflow and its components, as well as performs evaluation. Here we do not distinguish between these roles, referring to them collectively as *workflow developer*.

Moreover we could identify two additional roles. A *service developer* develops generic e-BioInfra services and/or integrates a workflow into the e-BioInfraGateway, and the *operations support* team is responsible for keeping the platform running and providing user support, for example, communication with the BiG Grid support team. Finally the workflows or the integrated experiments are executed by a *biomedical researcher* to perform data analysis (called clinical user in [21]) .

2.3. Usage Patterns and User Profiles

We collected information about the e-BioInfra to identify the roles taken by each user and the interfaces they use most regularly. We also determined whether or not a user has a grid certificate, which is normally an indication of user expertise. From a total number of 40 registered users, seven are inactive and seven are only interested in accessing data in a particular project, so they were not considered here. Table 1 summarises the data collected for the remaining 26 users, and figure 4 summarises the totals for each role that adopt a given type of user interface. Note that biomedical researchers prefer the web interface, whereas the developers and support team prefer the CLI and VBrowser interfaces. The VBrowser is widely used by all groups. Although the usage results show that most VBrowser users are also CLI users, this should not be interpreted as disadvantage of any over the other. All of the interfaces are complementary and the users choose an interface to the e-BioInfra based on their preference and requirements. For example, e-BioInfraGateway is useful for highly-demanded well-established applications, VBrowser is more friendly for occasional usage, and CLIs are better for repetitive tasks.

Although there is no clear distinction between user profiles, we observed five major classes within a wide spectrum of expertise and interest:

(i) users with biomedical background who just want to perform data analysis using existing tools. These are usually not interested in technical details, being most familiar with GUIs. They usually take only the biomedical researcher role, preferring a web interface and avoiding use of grid certificates;
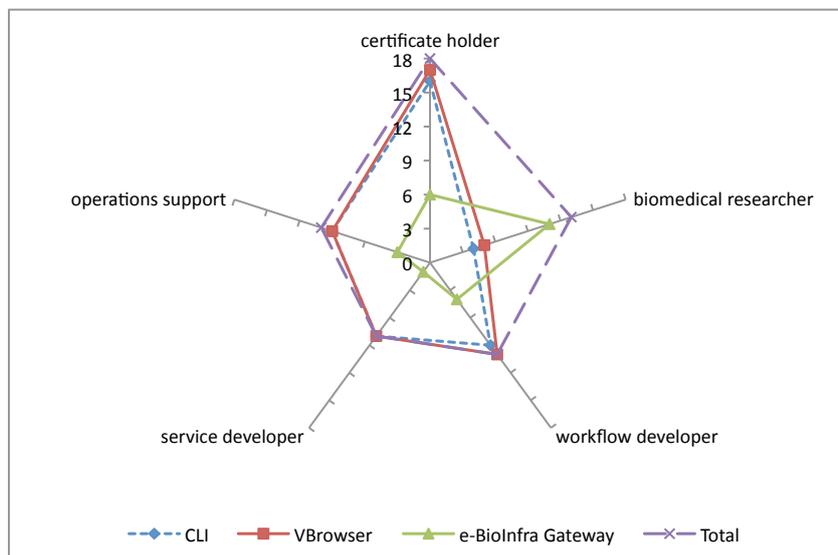
7

Figure 4. The roles of the users categorized by the type of user interface they regularly use.

Table 1. Roles taken and types of interfaces used by the e-BioInfra users. See text for user profile.

user	☺☺☺	☺☺☺☺	☺☺☺☺☺☺	☺☺☺☺☺	☺☺☺☺☺☺☺☺	total 26	%
certificate holder	✓✓✓	✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓		18	69
biomedical researcher				✓✓✓✓✓	✓✓✓✓✓✓✓✓	13	50
workflow developer			✓✓✓✓✓✓	✓✓✓✓		10	38
service developer		✓✓✓✓	✓✓✓	✓		8	31
operation support	✓✓✓	✓✓✓	✓✓ ✓✓			10	38
CLI user	✓✓✓	✓✓✓✓	✓✓✓✓ ✓	✓✓✓✓		16	62
VBrowser user	✓✓	✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓		17	65
e-BioInfraGateway user		✓	✓✓	✓✓✓	✓✓✓✓✓✓✓✓	14	54
user profile	(v)	(iv)	(iii)	(ii)	(i)		

(ii) users with biomedical background who are also interested in technical details and try to improve their data analysis by developing new methods and workflows. They have programming skills and do not mind using CLIs, but they prefer interacting with the system through a GUI. They usually take both the biomedical researcher and workflow developer roles;

(iii) users that only develop workflows (workflow developer role). These usually have technical expertise, but not necessarily a biomedical background. The VBrowser and CLI are the preferred interfaces.

(iv) programmers and software engineers who usually have the informatics background and take the middleware service developer role. CLI in this case is unavoidable for low-level service development and debugging; and

(v) system administrators who are familiar with Unix-based operating systems and comfortable with command-line interfaces. They usually take the support role as external parties (e.g., operators of the BiG Grid facilities).

We also analysed the activity associated with user profiles and roles. Because of the distributed, loosely coupled and layered architecture of the grid, it is rather difficult or even impossible to collect complete usage information per user. For instance, the workflow service collects the number of workflows per user, but not the CPU usage. Although it is in principle possible to extract this usage from the DIANE pilot job framework, the information is not complete and comparable because not every user uses DIANE. Moreover, some sites publish grid activity only in anonymized way and/or for coarse-grained granularity (i.e., per VO), so the accounting information available for us is not reliable for this purpose. As an indication of activity we therefore only present the number of executed workflows. Table 2 summarizes the total number of workflows submitted by the users in the first five months of 2011, when all of the three user interfaces were operational. The numbers are categorized based on two orthogonal qualities: the interface used to submit them (web and non-web) and whether the activity refers to "real" data analysis or to development and support. As expected, note that the number of workflows submitted via the web interface is larger for real data analysis, whereas the CLI and VBrowser are more used for development and support. Also note that the number of workflows for both types of activities are rather similar, although we expect that the workflows executed for data analysis are larger and run longer. Unfortunately we lack data to confirm this hypothesis.

Table 2. Number of workflows submitted by the e-BioInfra users.

	data analysis	development	total
submitted through CLI or VBrowser	1450	1350	2800
submitted through the e-BioInfraGateway	566	84	650
total	2016	1434	3450

3. Related Work

The same basic set-up (MOTEUR web-service and VBrowser interface) is adopted by GATE-Lab project, which enables running GATE [10] on several computing platforms. It splits the simulation into sub-tasks that are submitted to the grid, monitors the simulation until completion, retrieves and merges the outputs into a location accessible with a simple web URL, and keeps track of the simulations history [5].

Web interfaces are common in a large number of projects providing access to grid resources. The EGI User Support website (VRC)[7] provides a list of discipline-specific gateways to enable researchers to operate their data analysis in a manner that is more closely aligned to their own particular domain-specific skill sets. For example, the eNMR and WeNMR projects created an e-infrastructure for the biomolecular nuclear magnetic resonance (NMR) user community. The platform integrates and streamlines the compu-

---

[7]https://www.egi.eu/user-support/vrc_gateways/

tational approaches necessary for NMR data analysis and structural modelling. Access to the e-infrastructure is provided through a portal integrating commonly used NMR software and grid technology [2].

Portal frameworks compatible with the Java Portlet APIs (JSR-168 and JSR-286) are becoming popular to build user interfaces to grid-enabled systems. GridSphere is a common portal solution used in grid computing environment [17] that provides a framework compatible with JSR-168. Portlets have a standard API and provide a model for developing specific components for each application; core portlets are available e.g., for authentication. GridSphere is used in many grid community projects, e.g., the MediGRID project [12], the ViroLab project [3]. The P-GRADE portal supports creation, execution, and management of traditional and parameter study grid workflows using a variety of middleware, e.g., gLite and Globus. The first generation of P-GRADE adopts GridSphere as framework [6], but the new generation, coined WS-PGRADE, adopts the Liferay portal framework[8]. P-GRADE is the basis of the simulation platform under development by the SHIWA project[9], which will enable the execution of existing workflows on distributed computing infrastructures using various workflow engines (e.g., ASKALON, Pegasus, P-Grade, MOTEUR, Triana, and GWES). As part of the SHIWA Simulation Platform, the SHIWA Repository facilitates publishing and sharing workflows, and the SHIWA Portal enables their actual enactment. GENIUS (Grid Enabled web eNvironment for site Independent User job Submission) is an application independent web-based portal dedicated to EGEE gLite infrastructure [1]. It provides general security, job submission and data management services. The portal is implemented using another framework, EnginFrame[10]. The Liferay-based MoSGrid (Molecular Simulation Grid) portal provide grid services for performing molecular simulations on the D-Grid infrastructure, also including services for the annotation of results and data mining [8].

4. Conclusion and Future Work

In the past six years, the e-BioInfra evolved to facilitate access to grid resources for biomedical researchers. To support a wide spectrum of user profiles, with different expertise and requirements, the platform now offers a variety of user interfaces. Users choose which interface to use based on the role they take and their preference. We observed that an easy to use web interface is more popular among biomedical researchers, whereas CLIs and desktop applications are more used for development and support.

We showed that the platform is used by people within a wide spectrum of expertise and interest, for service development, workflow development and biomedical data analysis. The service oriented architecture of the platform enables it to be flexible and extensible, as we observed its improvement at the same time as it was used in production. In addition we observed that a fruitful e-science biomedical research community consists of people with a variety of expertise, from informatics background to biology, where some provide operations support such that others can focus on biomedical data analysis.

The e-BioInfra is not complete and there is still much room for improvement. Our vision is to enhance the web interface to enable collaboration between researchers and

---

[8]http://www.liferay.com/
[9]http://shiwa-workflow.eu
[10]http://www.enginframe.com

facilitate access to grid resources more easily and dynamically. We plan to adopt the Liferay portal framework, and we expect to reuse a major portion of the e-BioInfraGateway code when migrating to it. Finally, today the user interfaces are disconnected. Finding out about how users choose for one or another will enable us to instrument them to be aware of each other and facilitate navigation between the various types of interfaces.

References

[1] R. Barbera, A. Falzone, V. Ardizzone, and D. Scardaci. The GENIUS Grid Portal: Its Architecture, Improvements of Features, and New Implementations about Authentication and Authorization. *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2007. WETICE 2007. 16th IEEE International Workshops on*, pages 279–283, 2007.

[2] A. M. J. J. Bonvin, A. Rosato, and T. A. Wassenaar. The eNMR platform for structural biology. *Journal of structural and functional genomics*, 11(1):1–8, mar 2010.

[3] M. Bubak, T. Gubala, M. Malawski, B. Balis, W. Funika, T. Bartynski, E. Ciepiela, D. Harezlak, M. Kasztelnik, J. Kocot, D. Krol, P. Nowakowski, M. Pelczar, J. Wach, M. Assel, and A. Tirado-Ramos. Virtual Laboratory for Development and Execution of Biomedical Collaborative Applications. In *2008 21st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 373–378. IEEE, 2008.

[4] M. Caan, F. Vos, A. van Kampen, S. Olabarriaga, and L. van Vliet. Gridifying a Diffusion Tensor Imaging Analysis Pipeline. *Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on*, pages 733–738, 2010.

[5] S. Camarasu-Pop, T. Glatard, J. T. Moscicki, H. Benoit-Cattin, and D. Sarrut. Dynamic Partitioning of GATE Monte-Carlo Simulations on EGEE. *Journal of Grid Computing*, 8(2):241–259, 2010.

[6] Z. Farkas and P. Kacsuk. P-GRADE Portal: A generic workflow system to support user communities. *Future Generation Computer Systems*, 27(5):454–465, may 2011.

[7] B. Fischl, A. van der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, and A. M. Dale. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1):11–22, 2004.

[8] S. Gesing, I. Marton, G. Birkenheuer, B. Schuller, R. Grunzke, J. Krüger, S. Breuers, D. Blunk, G. Fels, L. Packschies, A. Brinkmann, O. Kohlbacher, and M. Kozlovszky. Workflow Interoperability in a Grid Portal for Molecular Simulations. In G. A. R Barbera and G. La Rocca, editors, *Proceedings of the International Worshop on Science Gateways (IWSG2010)*, pages 44–48, Consorzio COMETA, 2010. Consorzio COMETA.

[9] T. Glatard, J. Montagnat, D. Lingrand, and X. Pennec. Flexible and Efficient Workflow Deployment of Data-Intensive Applications On Grids With MOTEUR. *International Journal of High Performance Computing Applications*, 22(3):347–360, aug 2008.

[10] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assie, D. Autret, S. Avner, R. Barbier, M. Bardies, P. M. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A. F. Chatziioannou, Y. Choi, Y. H. Chung, C. Comtat, D. Donnarieix, L. Ferrer, S. J. Glick, C. J. Groiselle, D. Guez, P. F. Honore, S. Kerhoas-Cavata, A. S. Kirov, V. Kohli, M. Koole, M. Krieguer, D. J. van der Laan, F. Lamare, G. Largeron, C. Lartizien, D. Lazaro, M. C. Maas, L. Maigne, F. Mayet, F. Melot, C. Merheb, E. Pennacchio, J. Perez, U. Pietrzyk, F. R. Rannou, M. Rey, D. R. Schaart, C. R. Schmidtlein, L. Simon, T. Y. Song, J. M. Vieira, D. Visvikis, R. Van de Walle, E. Wieers, and C. Morel. GATE : a simulation toolkit for PET and SPECT. *arXiv.org*, physics.med-ph(19):4543–4561, aug 2004.

[11] P. L. Klarenbeek, P. P. Tak, B. D. C. van Schaik, A. H. Zwinderman, M. E. Jakobs, Z. Zhang, A. H. C. van Kampen, R. A. W. van Lier, F. Baas, and N. de Vries. Human T-cell memory consists mainly of unexpanded clones. *Immunology Letters*, 133(1):42–48, sep 2010.

[12] D. Krefting, J. Bart, K. Beronov, O. Dzhimova, J. Falkner, M. Hartung, A. Hoheisel, T. A. Knoch, T. Lingner, Y. Mohammed, K. Peter, E. Rahm, U. Sax, D. Sommerfeld, T. Steinke, T. Tolxdorff, M. Vossberg, F. Viezens, and A. Weisbecker. MediGRID: Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems-The International Journal Of Grid Computing Theory Methods And Applications*, 25(3):326–336, 2009.

[13] E. Laure, F. Hemmer, F. Prelz, S. Beco, S. Fisher, M. Livny, L. Guy, M. Barroso, P. Buncic, P. Z. Kunszt, A. Di Meglio, A. Aimar, A. Edlund, D. Groep, F. Pacini, M. Sgaravatto, and O. Mulmo. Middleware for the next generation Grid infrastructure. (EGEE-PUB-2004-002):4 p, 2004.

[14] A. Luyf, B. van Schaik, M. de Vries, F. Baas, A. van Kampen, and S. Olabarriaga. Initial steps towards a production platform for DNA sequence analysis on the grid. *BMC Bioinformatics*, 11(1):598, 2010.

[15] J. Montagnat, B. Isnard, T. Glatard, K. Maheshwari, and M. Fornarino. A data-driven workflow language for grids based on array programming principles. *WORKS '09: Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science*, nov 2009.

[16] J. T. Moscicki, J. T. Moscicki, M. Lamanna, M. Lamanna, M. Bubak, M. Bubak, P. M. A. Sloot, and P. M. A. Sloot. Processing moldable tasks on the grid: Late job binding with lightweight user-level overlay. *Future Generation Computer Systems*, 27(6):725–736, jun 2011.

[17] J. Novotny, M. Russell, and O. Wehrens. Proceedings. 30th Euromicro Conference, 2004. In *Proceedings. 30th Euromicro Conference, 2004.*, pages 412–419. IEEE, 2004.

[18] S. Olabarriaga, P. De boer, K. Maheshwari, A. Belloum, J. Snel, A. Nederveen, and M. Bouwhuis. Virtual Lab for fMRI: Bridging the Usability Gap. In *e-Science and Grid Computing, 2006. e-Science '06. Second IEEE International Conference on*, page 53, dec 2006.

[19] S. D. Olabarriaga, T. Glatard, K. Boulebiar, and P. T. de Boer. From "low hanging" to "user ready": initial steps into a HealthGrid. *Studies in health technology and informatics*, 138:70–79, 2008.

[20] S. D. Olabarriaga, T. Glatard, and P. T. de Boer. A Virtual Laboratory for Medical Image Analysis. *Ieee Transactions on Information Technology in Biomedicine*, 14(4):979–985, 2010.

[21] S. D. Olabarriaga, J. G. Snel, C. P. Botha, and R. G. Belleman. Integrated Support for Medical Image Analysis Methods: From Development to Clinical Application. *Information Technology in Biomedicine, IEEE Transactions on*, 11(1):47–57, 2007.