

SPECIAL ISSUE PAPER

A data-centric neuroscience gateway: design, implementation, and experiences

Shayan Shahand^{1,*}, Ammar Benabdelkader¹, Mohammad Mahdi Jaghoori¹,
Mostapha al Mourabit¹, Jordi Huguet², Matthan W.A. Caan²,
Antoine H.C. van Kampen^{1,3} and Sílvia D. Olabarriaga^{1,*}

¹*Bioinformatics Laboratory, Dept. of Clinical Epidemiology Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, The Netherlands*

²*Brain Imaging Center, Dept. of Radiology, Academic Medical Center, University of Amsterdam, The Netherlands*

³*Biosystems Data Analysis, Swammerdam Institute for Life Science, University of Amsterdam, The Netherlands*

SUMMARY

Science gateways provide UIs and high-level services to access and manage applications and data collections on distributed resources. They facilitate users to perform data analysis on distributed computing infrastructures without getting involved into the technical details. The e-BioInfra Gateway is a science gateway for biomedical data analysis on a national grid infrastructure, which has been successfully adopted for neuroscience research. This paper describes the motivation, requirements, and design of a new generation of e-BioInfra Gateway, which is based on the grid and cloud user support environment (also known as WS-PGRADE/gUSE framework) and supports heterogeneous infrastructures. The new gateway has been designed to have additional data and meta-data management facilities to access and manage (biomedical) data servers, and to provide data-centric user interaction. We have implemented and deployed the new gateway for the computational neuroscience research community of the Academic Medical Center of the University of Amsterdam. This paper presents the system architecture of the new gateway, highlights the improvements that have been achieved, discusses the choices that we have made, and reflects on those based on initial user feedback. Copyright © 2014 John Wiley & Sons, Ltd.

Received 8 November 2013; Revised 10 February 2014; Accepted 17 March 2014

KEY WORDS: science gateway (SG); e-Science; computational neuroscience; medical image analysis; grid computing; virtual laboratory (VL); problem solving environment (PSE); virtual research environment (VRE)

1. INTRODUCTION

Science gateways (SGs) support scientists in e-Science endeavors. De Roure et al. [1] described the requirements of e-Science environments as a spectrum with two ends. One end is characterized by automation, virtual organizations of services, and the digital world, and the other end is characterized by interaction, virtual organizations of people, and the physical world. Orthogonal to these requirements at both ends is the issue of scale, for example, of virtual organizations, computation, storage, and the complexity of relationships between them. Increasing scale demands automation and, as highlighted by Hey and Trefethen [2], computer scientists have the research challenge of creating high-level intelligent services that genuinely support e-Science applications. Such services, for example, SGs, should go beyond straightforward access to computing resources, and also include

*Correspondence to: Shayan Shahand and Sílvia D. Olabarriaga, Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center of University of Amsterdam, The Netherlands.

†E-mail: s.shahand@amc.uva.nl; s.d.olabarriaga@amc.uva.nl

support to construct and manage virtual organizations, as well as to manage the scientific data deluge in the scholarly cycle including hypothesis, experimentation, analysis, publication, research, and learning.

A large number of communities are therefore facing the challenge of building SGs. A recent collaboration resulted in the European Grid Infrastructure (EGI) SG Primer [3], where issues involved in SG design, implementation, and operation are presented and discussed. According to this Primer, SGs are desktop or Web-based interfaces to a set of applications and data collections. SGs comprise front-end and back-end components, and they offer services that facilitate access to computing and storage resources, as well as services provided by distributed computing infrastructures (DCIs). Moreover, SGs support collaboration between researchers through exchange of ideas, tools, and datasets. From a functional perspective, there are *SG frameworks* and *SG instances*. SG frameworks implement generic functionalities such as security, workflow and data management, and DCI access; examples are grid and cloud user support environment (also known as WS-PGRADE/gUSE) [4], Catania SG [5], Hubzero [6], and DARE [7] frameworks. SG instances are community-specific SGs, with tailored interfaces and services for a specific application domain. SG instances can be built using SG frameworks or with custom software stacks. There are currently a large number of SG instances, which will be discussed in more details in Section 6.

The e-BioInfra Gateway [8] was an SG instance for analysis of large scale biomedical data on the Dutch e-Science Grid [9]. It was designed to simplify usage of this infrastructure by biomedical researchers by providing services such as community grid certificate and semi-automatic file transport to and from the grid resources. It was implemented based on a custom framework and deployed at the Academic Medical Center (AMC) of the University of Amsterdam, the Netherlands. Since its deployment in production (early 2011), researchers have successfully performed large computations on the Dutch grid infrastructure via the e-BioInfra Gateway with minor help from the support team. For example, Peters et al. [10], Wingen et al. [11], Rienstra et al. [12], and de Kwaastieniet et al. [13] have already published results of their neuroscience research based on the data analysis performed via the e-BioInfra Gateway.

The gateway structured the system information and allowed for extensions with new data analysis methods. This enabled (external) developers to extend it with 10 applications, six for medical imaging, three for next generation sequencing data analysis, and one as part of the mass-spectrometry tool-chain. Thirty users utilized the e-BioInfra Gateway, most of which were from the Brain Imaging Center of the AMC (BIC [14]).

Although the e-BioInfra Gateway can be considered a success story, our experience indicated the need for further improvement in particular regarding support for richer data resources and operations. We therefore designed, implemented, and tuned a new gateway specifically for the computational neuroscience research community of AMC. The most important new features of the new gateway include data and information management, support for meta-data that is used and generated during the execution of complex data processing, and additional functionalities for system operation and administration.

In this paper, we discuss the motivation for a new gateway in Section 2 and describe a requirement analysis for it in Section 3. The system architecture is explained in Section 4. Initial user feedback is presented in Section 5, followed by comments about related work in Section 6. Finally, discussion and conclusions are presented in Sections 7 and 8.

2. MOTIVATION FOR A NEW GATEWAY

In this section, the functionalities of the previous gateway are briefly explained, and the experiences gained during its operation are reflected upon. These experiences motivated the design of the next generation of the e-BioInfra Gateway.

2.1. The previous gateway

A detailed description of the previous e-BioInfra Gateway can be found in [8]. In a nutshell, the user could authenticate with username and password, select the application to run, select the input files

and other parameters, and start a so called *experiment*. Then he/she could monitor the experiment and, when finished, retrieve the results.

The processing on grid resources was handled by the MOTEUR [15] workflow management system (WfMS) and the DIANE pilot job framework [16]. Minimal provenance information was recorded, namely, the names of the files used in an experiment and the history of changes in an experiment status. More detailed provenance information about workflow execution was collected and displayed by a separate system after the processing was completed [17].

Because medical imaging data files are typically large, their transport was not performed directly via the e-BioInfra Gateway web interface, but via an FTP server that was located in the trusted network of the hospital and synchronized automatically with a directory on the grid storage resources. Therefore, the user had to upload the data to the server before performing the steps in the preceding texts, and could retrieve the results from the same FTP server when the experiment was completed.

2.2. Reflections about the previous gateway

The functionalities offered by the previous gateway mainly included the following: transparent authentication and authorization with grid resources using a robot certificate; semi-automatic data transfer between gateway and grid storage workflow processing management, including logging and monitoring; and an extensible set of applications for various biomedical domains. These functionalities were mainly organized around applications, underlying resources, and their frameworks.

In almost 2 years of experience with gateway operation and user support, we faced challenges discussed in the succeeding texts that made us realize that the gateway should be organized around *data* instead of *applications*.

In the previous gateway, a large number of errors were caused by invalid input data, which are referred to as *data-related* errors. Users typically had difficulty to prepare files for processing with the gateway applications, which involved the steps for file (re-)formatting, naming, and transport. Also, they were not aware of the types of data that can be processed by each application. Although the data-related errors were significantly reduced after training or reading the user manual, we realized that the data preparation and transport process should be improved with further automation.

Whenever the errors were not data-related, they were mostly related to the changes or maintenance operations performed on the grid infrastructure, which are referred to as *computing-related* errors. Exposing users to the computing-related errors turned out to be both unnecessary and overwhelming for the users. A system administrator usually could fix those errors by simple actions such as resource blacklisting and resubmitting the failed experiment. However, it was not straightforward in the old gateway to resubmit parts of the experiments on behalf of the users.

Another necessary improvement was motivated by the evolution in the computing infrastructure. Originally, the e-BioInfra Gateway was meant to facilitate access to grid resources. However, in the past years, other resources have become available for research, such as local clusters at the AMC and a national high-performance cloud. Another solution was required to exploit these additional resources.

Finally, the need for adopting a more sustainable software stack was evident. Although our custom framework fulfilled the needs at first, as a small research group, it was difficult to maintain and extend it. In particular, keeping up with all the developments related to DCIs requires significant effort and expertise that can be achieved by utilizing an SG framework.

2.3. Preconditions for a new gateway

Recently, the neuroscience research community of AMC has decided to adopt a data server for their research scans. The data are generated by the scanner and directly imported into the data server, which keeps both the raw data and the meta-data. To facilitate research data processing, the Radiology department decided that this research data server could be connected to the gateway.

Because of security regulations for processing medical research data, the data server is hosted inside the AMC firewall. The in-house computing clusters and national grid computing resources that are used for data processing are located inside and outside the AMC network, respectively. The

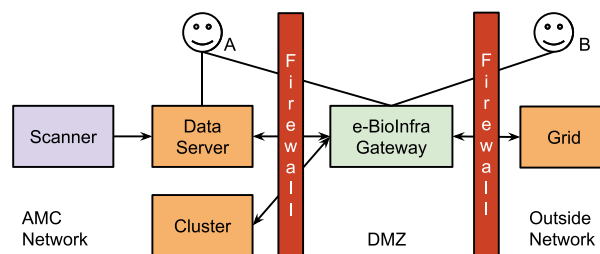


Figure 1. The resources related to the e-BioInfra Gateway for computational neuroscience and their network location: inside or outside the AMC network, or at the demilitarized zone. User *A* is inside the network and can access the data directly. User *B* is outside of the network and therefore only has access to the gateway and limited meta-data.

e-BioInfra Gateway server is located in the demilitarized zone (DMZ) of the AMC network, which means that only some of the gateway services are visible from outside the network and, similarly, only some of the data server and in-house computing cluster services are visible from the demilitarized zone. Figure 1 illustrates the resources related to the e-BioInfra Gateway for computational neuroscience and their network location.

The envisioned usage scenario for this system is as follows. The users inside the AMC network can import their data into the data server either directly from the scanner or by uploading it to the data server. The data are automatically preprocessed according to pre-defined rules, for example, it is pseudonymised and converted to a more compact format, and its meta-data is extracted from the Digital Imaging and COmmunication in Medicine headers [18]. The users, both inside and outside the AMC network, are able to query and filter the data based on its meta-data and initiate and monitor data processing, regardless of their location. After the data processing is completed, users should be able to download the results from the data server only if they are inside the AMC network. The system administrators, who are also located inside the AMC network, should be able to monitor all data processing activities and inspect them in more detail if any error happens.

3. REQUIREMENTS ANALYSIS

In a previous work, we described in detail the typical phases of computational neuroscience studies in [19], namely, *study design, data acquisition, data handling, processing, analysis, and publication*. Based on the analysis of these phases, the actors who are involved in each phase, and the tasks that they perform, in that paper, the properties and functionalities of SGs to support computational neuroscience research communities are identified. In summary, the required properties and functionalities include the following: sharing of data and methodology; satisfying security and privacy regulations; scalable, transparent, and flexible management of storage and computing resources; literature discovery; collaboration support; meta-data, data, workflow, and provenance management; and visualization.

The design of the new gateway presented in this paper takes into account these desired properties, as well as the experiences and preconditions presented in Section 2. In particular, we focused on additional functionalities that would put *data* at the center of interaction between the user and the gateway. A *data-centric* gateway should provide necessary tools and services to the users to interact with their data, for example, for data discovery, exploration, preparation, and processing.

The following functionalities should be provided by the new data-centric gateway:

- (i) Unified, secure, and easy access to data and related meta-data stored on distributed and heterogeneous data servers. Users should be able to transparently query, explore, process, and analyze data from a single interface, without bothering about the data location or format, or how the data are retrieved and transported for further processing.
- (ii) Automatic and interoperable file transport and processing on different infrastructures (e.g., data servers, cluster, and grid). Low-level technical details should be hidden from the

users, such as different communication protocols, middleware services, and authorization mechanisms.

- (iii) Assistance for users to choose the correct data processing method based on meta-data.
- (iv) Automatic provenance information collection about the methods, parameters, and input files used for processing. This provenance information can be used in troubleshooting, to track the data lineage, and for statistics.
- (v) Single sign-on facility to authenticate and authorize transparently to various computing and storage resources using user or community credentials.
- (vi) Streamlined operations of the gateway by its system administrators. They should be able to access log files easily, communicate the causes of errors with the users, and restart the faulty data processing on their behalf.

In addition to these functionalities, the new gateway should be

- (i) *extensible*, to easily connect to new data or compute resources, and accommodate new data types, applications, and user groups;
- (ii) *customizable*, to support preferences and configurations for both end-users and system administrators;
- (iii) *scalable*, to gracefully support the growth of user community and its needs for resources, as well as infrastructures capacity and heterogeneity; and
- (iv) *sustainable*, to be able to maintain the gateway software with minimal costs, while its underlying infrastructure changes.

4. SYSTEM DESIGN AND IMPLEMENTATION

Figure 2 illustrates the layered architecture of the new e-BioInfra Gateway. At the bottom, the Resource layer (dark orange) with several DCIs (i.e., local clusters and grid) and data resources (i.e., Radiology research data server). These resources are utilized through middleware services contained in the second layer (light orange). High-level services contained in the third layer (blue) provide an abstraction to interact with the middleware, such as workflow management and data transport. Finally, the Presentation layer (green) contains the interfaces for user interaction. The two topmost layers (green and blue) are implemented using generic SG framework components provided by

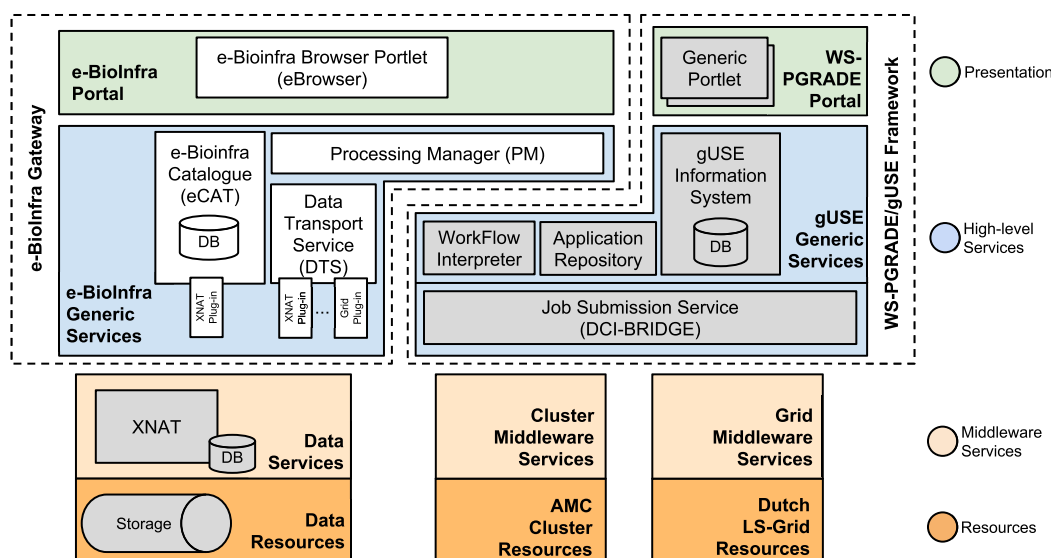


Figure 2. Layered architecture of the e-BioInfra Gateway based on the WS-PGRADE/grid User Support Environment Science gateway framework. Gray boxes represent existing third-party components, and white boxes denote new components. See text for more details.

WS-PGRADE/gUSE (at the right), and components developed for the new gateway (at the left). The components of the new gateway complement the functionality of WS-PGRADE/gUSE framework for the specific case of the e-BioInfra Gateway for computational neuroscience.

The core of the new e-BioInfra Gateway is made of the following components: e-BioInfra Catalogue (eCAT), data transport service (DTS), processing manager (PM), and e-BioInfra Browser portlet (eBrowser). They are loosely coupled and communicate via well-defined APIs, an approach that paves the road towards a service-oriented architecture and facilitates their reuse to build other gateways for different scientific applications. These components also utilize the API of the WS-PGRADE/gUSE components to implement the functionalities of the e-BioInfra Gateway for computational neuroscience.

Below the components that are more relevant for a data-centric SG, namely, data services and the core components (white boxes in Figure 2) are presented in further detail. For completeness, the WS-PGRADE/gUSE SG framework is also introduced briefly. Finally, the interactions between these components are illustrated based on a use case.

4.1. e-BioInfra Catalogue

The eCAT has been designed to facilitate the data and meta-data management functionalities at the gateway. It is a central store for user and system-level information that uses and implements a data model with the following main entities: User, Project, Data, Meta-data, Resource, Credential, Application, Processing, Submission, and Submission Status. The main relationships between these entities are illustrated in Figure 3.

In the eCAT data model (Figure 3), a User participates in Projects, which provides the scope for access control. Data entities are included in, and are processed within, the scope of Project entities. Each User has one or many Credentials that are used by the gateway to access resource(s) on the user's behalf. A Resource can be a computing resource, a storage resource, or both. A data server is an example of a data resource, and grid or clusters are examples of resources both for data and computing. Each Data has at least one replica on a data resource and has meta-data attached to it. Meta-data is represented by a key-value pair. Users have also access to Applications consisting of validated and ready-to-use workflows that wrap some legacy code for data analysis. Applications have inputs and generate outputs; they also have affinities with particular data types and formats. The outputs of applications are also stored as Data entities that are annotated with provenance information about the applications that generated them. When a user processes a certain Data with a specific Application, the information about this activity is captured by eCAT as a Processing entity. Each Processing includes one or more workflow Submissions, depending on the cardinality of input data. A workflow is executed on a computing Resource. The provenance

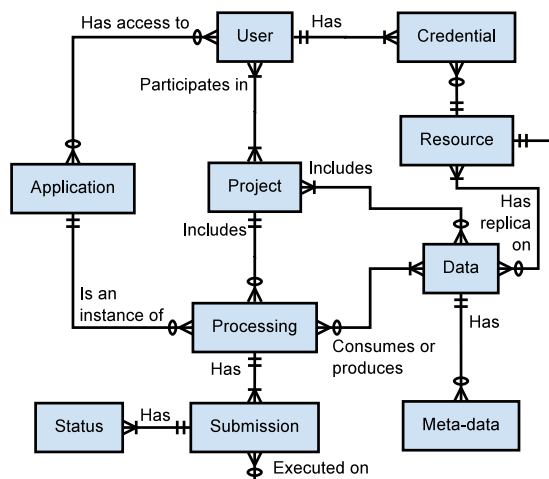


Figure 3. Simplified entity-relationship model of the information stored in the e-BioInfra Catalogue.

information about the Data consumed and produced during a Processing, the parameters, and the history of Submission Status are also stored in the eCAT database as relationships and attributes of these entities.

Note that eCAT is not meant to duplicate meta-data that is already stored on data servers; instead, it only stores pointers to information on the data servers. The only exceptions are some types of meta-data that are specific to user activities on the gateway, and which are not possible, nor of direct interest of research communities, to store in their data servers. eCAT retrieves and stores meta-data on heterogeneous data servers through *Plug-ins*, which are software modules attached to eCAT to enable programmatic communications with a specific data server.

4.2. Data servers

A data server can be as simple as an FTP server that contains the data in a hierarchy of directories. However, management of biomedical research data, with its growing size and complexity, requires domain-specific *Information Management Systems (IMSs)* with structured meta-data. There are several IMSs for management of biomedical research data and meta-data, electronic data exchange, archival, and security, and the research communities start to adopt such systems routinely. Additionally, every community has its own procedure to implement rules and regulations regarding the protection of biomedical research data, as well as policies for data sharing and archiving. Therefore, instead of replicating such efforts, we decided to rely on existing, external, biomedical research data and meta-data resources, as well as on their own security mechanisms and policies. In this way, the research community itself provides and manages the IMS, defining data ownership, access policies, and regulating data confidentiality and data privacy.

A popular IMS for medical imaging data and meta-data is the eXtensible Neuroimaging Archive Toolkit (XNAT) [20]. XNAT is an open-source IMS that offers an integrated framework for storage, management, electronic exchange, and consumption of medical imaging data and its complementary meta-data. XNAT provides a rich communication layer based on a RESTful API of resource-oriented web services. Because of these qualities, XNAT has been deployed at the Radiology department of AMC to implement a research data server. The XNAT server is connected to the e-BioInfra Gateway by agreement between the neuroscience community and the gateway providers. The data becomes available for processing at the gateway for authorized users only. Gateway users should provide their XNAT credentials before they are able to access data and meta-data on the XNAT. All of the API calls from the gateway to the XNAT are performed with user credentials.

EXtensible Neuroimaging Archive Toolkit implements an extensible data model that also has some fixed entities. In summary, XNAT ‘users’ have access to ‘projects’ that contain ‘subjects’ (i.e., people who have one or more scans) and their ‘image sessions’. Each image session includes one or more ‘scans’, and each scan has a many-to-many relationships with a specific entity called ‘reconstructed image’. A reconstruction image is the result of any processing software. The most relevant XNAT entities for our case are ‘projects’ and ‘scans’, which are mapped, respectively, to the `Project` and `Data` entities in the eCAT. The ‘reconstructed image’ entity of XNAT is used to store the processing provenance information for entities generated by the SG. In the new gateway, we developed an eCAT plug-in for XNAT using its RESTful API. This plug-in maps the eCAT data model into the XNAT data model by generating queries and parsing responses between them.

4.3. Data transport service

The DTS transports data between data servers and storage resources on DCIs. This service contacts the eCAT to retrieve information needed to access the data resources, such as hostname, type of resource, and credentials. From this information, it determines how to authenticate with the data server on behalf of the user, as well as how to authenticate with the storage resources of the DCI, typically with community credentials. It autonomously performs the data transfer using third-party mechanisms as much as possible to avoid bottlenecks. Similarly to the eCAT, DTS also has *Plug-ins* that implement the necessary functionalities to enable data transfer between resources. Currently, plug-ins are available for the XNAT data server and the gLite grid middleware. If some data have

been replicated on a DCI, the location of that replica is stored in the eCAT, and it can be retrieved later to avoid transporting the file again in the future.

4.4. WS-PGRADE/gUSE SG framework

WS-PGRADE/gUSE SG framework [4] is an open-source, workflow-oriented, and service-oriented framework that facilitates development, execution, and monitoring of scientific workflows on DCIs. It comprises the WS-PGRADE portal and the gUSE services. WS-PGRADE is based on the Liferay portal framework, which provides rich facilities for community management and customizable UIs. gUSE provides high-level services to access various DCI resources. These qualities motivated the choice for this SG framework to implement our gateway.

The most relevant gUSE services used by our gateway are as follows:

- *Job submission service or DCI-BRIDGE*: provides flexible and versatile access to a large variety of DCIs such as grids, desktop grids, clusters, clouds, and service-based computational resources. It also handles authentication and authorization to the configured DCIs transparently.
- *Workflow interpreter*: parses workflows, submits jobs to the DCI-BRIDGE, and retrieves their status for monitoring and fault-tolerance.
- *Application repository*: stores ready-to-use tested and configured workflows. These workflows are exported to the application repository by workflow developers, from where they are imported into user space for execution.
- *gUSE information system*: stores configurations of gUSE services and workflow related information such as workflow executions and their jobs status.

Additional facilities offered by gUSE are also very important for the implementation of our SG. The first is support for community credentials (robot certificates). The other is functionality to pause and resume workflow execution, which is used by the administrator.

The WS-PGRADE/gUSE framework also provides two APIs to create SG instances. We used the *application specific module (ASM)* API to utilize gUSE services, more specifically the *application repository* and the *workflow interpreter*.

The WS-PGRADE portal also offers a set of generic portlets to interact with gUSE services via web-based GUIs. These portlets are only visible to the developers and administrators of the e-BioInfra Gateway for computational neuroscience.

See [4] for the complete description of WS-PGRADE/gUSE services and portlets.

At the time of writing, the WS-PGRADE/gUSE framework did not have any facility to connect to data servers. Moreover, its data transport facilities were also limited. The additional components described earlier are meant to bridge this gap.

4.5. Processing manager

The PM takes care of preparation, submission, and monitoring of data processing applications that are executed on a given set of input files. All the details needed to run the application are obtained by querying eCAT, such as the gUse workflow, DCI, as well as the input ports, the output ports, and the relationship between them.

The steps carried out for each processing started by the user are the following; they are collectively called a Submission in the eCAT data model. First, the PM instructs the DTS to transport input files from the data server to the storage resources of the DCI on which the data processing will be performed. Then, it imports the workflow from the gUSE application repository and configures it with the physical location of input data. All workflows are configured to run with community credentials, such as in the previous gateway. The configured workflow is submitted to and executed by the gUSE workflow interpreter. The workflow execution is monitored by the PM autonomously. When the workflow execution is completed successfully, the PM instructs the DTS to transport the results back to the data server.

Note that each processing started by the user can generate one or more submissions. This depends on the number the input data files and the relationship between input data and results for each application. In most applications, there is a one-to-one relationship between input data and results, that

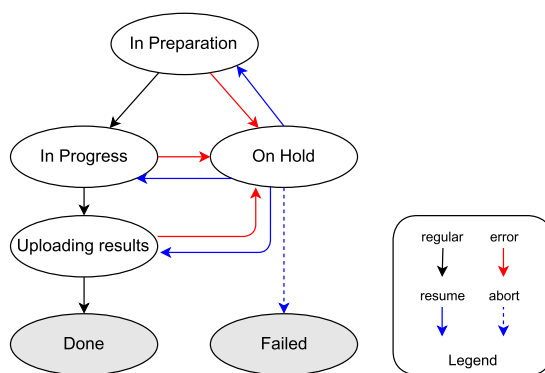


Figure 4. Submission state diagram. Red transitions occur on error and blue transitions occur by manual intervention of the administrator.

is, one result is generated for each input. In these cases, the processing consists of n workflow submissions, one for each of the n input data files. In other applications, a single result is generated for a collection of input data files, and therefore, a single workflow is submitted. Submitting one workflow for each processing result, instead of using parameter sweep capabilities of the WfMS, is motivated by the need for fine-grained control and monitoring of workflow execution. It also facilitates linking the results generated at the output ports of a workflow to its inputs, which is necessary for provenance collection. Note that the multiple workflow submissions for a processing are hidden from the user, whereas he/she can obtain progress information about the individual processing tasks transparently.

Each individual submission goes through the states illustrated in Figure 4, which correspond to status information shown to the users and system administrators. It is first in the *In Preparation* state, during which the input data is transferred from the data server to the target computing resource, and the workflow itself is imported, configured, and submitted through the gUSE ASM API. After successful submission, it reaches the *In Progress* state, during which the workflow is executed by the WfMS on the target computing resource. When the workflow execution completes successfully, the Submission moves to the *Uploading* state, during which the results are uploaded to the data server. Finally, when all the previous steps were completed successfully, the status changes to the *Done* state, and the results became available for the user via the interface. The user can also abort the submission at any time.

If any problem is detected during any of the operations performed for preparation, submission, workflow execution, and data transfers, the Submission moves to *On Hold* state, and a notification is sent to the system administrator. He/she then investigates and troubleshoots the error using information about that particular Submission that is presented on the administrator's dashboard (Figure 5). If the error can be fixed, the workflow is *resumed* and gUSE continues execution from the last successful job. This is often the case of errors related to the DCI, for example, a failed job or unavailable file. Otherwise, the administrator *aborts* the submission, which causes it to move into the *Failed* state. At this point, a message can be written to the user providing high-level information about the cause of the error and the actions to take. Typically, these are data-related errors, as all the DCI-related errors will be handled automatically by gUSE or manually by the administrator.

4.6. e-BioInfra Browser portlet

The eBrowser is part of the Presentation layer. It provides a web-based UI to interact with the e-BioInfra generic services. Instead of contacting the services directly, eBrowser retrieves information from eCAT to provide a homogeneous view to the users and system administrators to browse data, projects, and data processing instances. The eBrowser interfaces are adapted based on the roles that are assigned to the user profile: neuroscientist, called user here, and administrators.

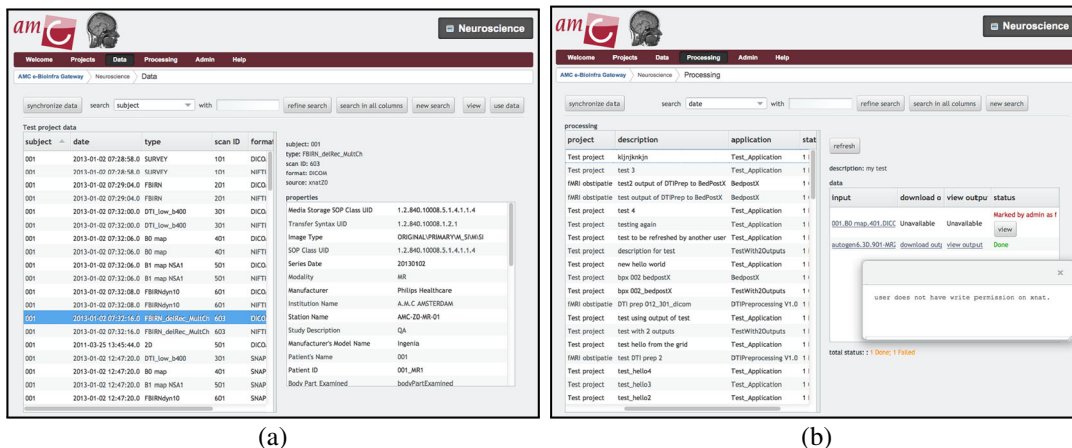


Figure 5. eBrowser examples for user view. (a) Data browsing: data list on the left and meta-data of selected item on the right. (b) Processing monitoring: list of processings on the right, progress information about selected processing on the left and box shows error message sent by administrator.

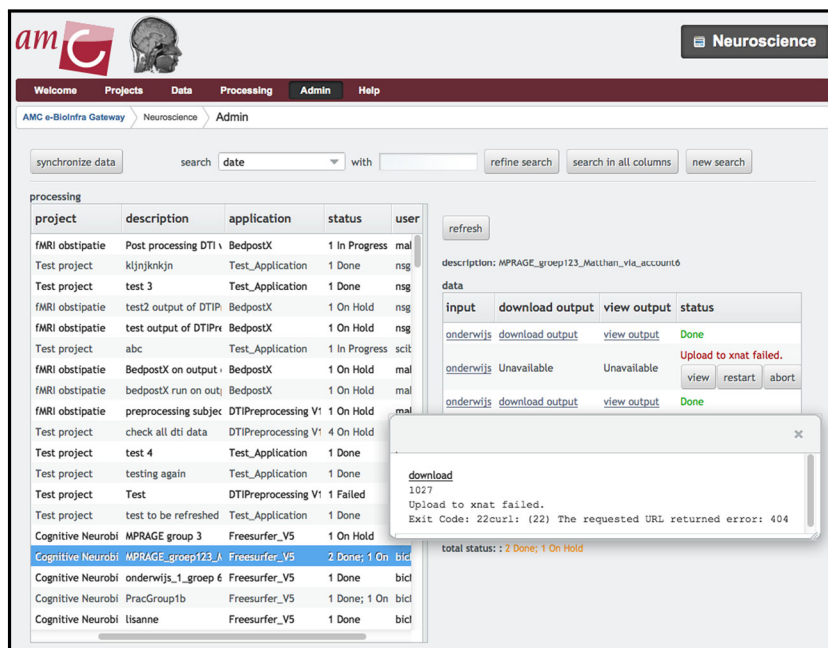


Figure 6. Administrator processing view: all processings on the right, detailed information about selected processing and its submissions on the left, and box shows detailed error message.

eBrowser essentially enables users to start, manage, and monitor data processing. Figure 5 depicts these UIs. When a user selects one or many data items to process, the eBrowser only displays the applications that are compatible with the selected data to the user based on the meta-data and application specifications.

The eBrowser also provides interfaces for system administrators. The administrator’s dashboard displays monitoring information about all of the user data processing activities and enables intervention on error. For example, in case of a failure during the execution of a workflow, a brief error message is displayed at the dashboard, and more details can be obtained by clicking the *View* button (Figure 6). The administrator can choose to *Resume* or *Abort* the execution of the workflow, and to send a high-level message for the user (e.g., the input file is corrupted).

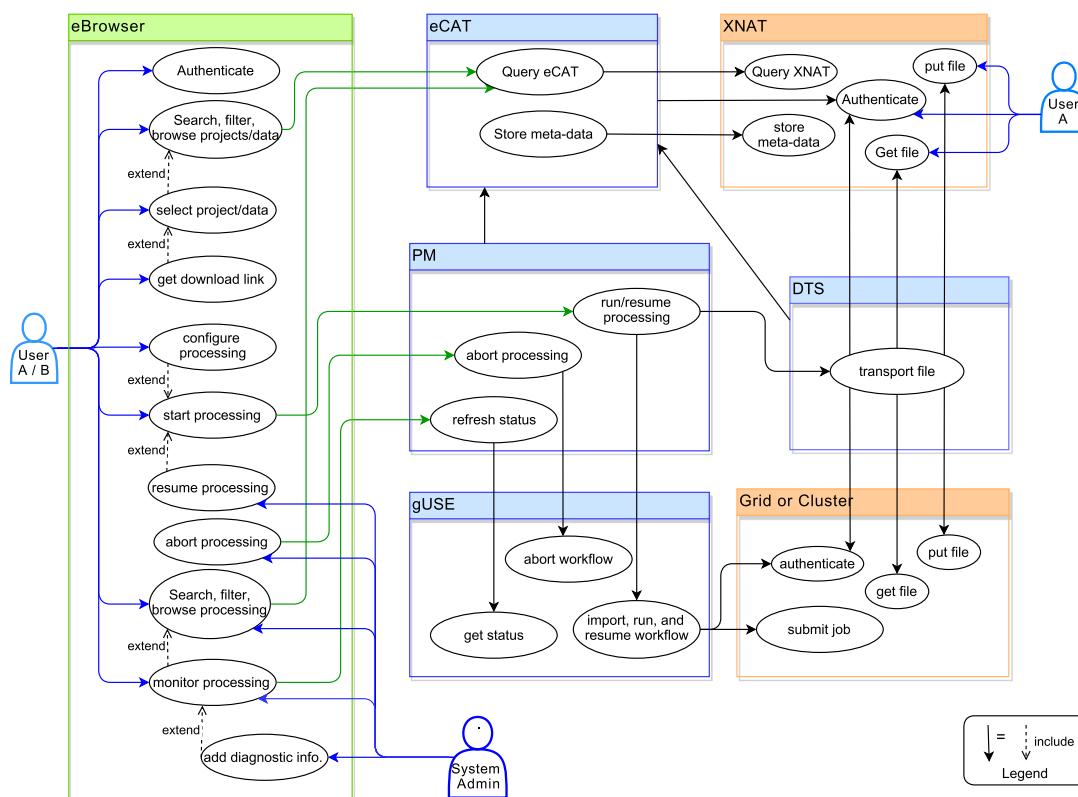


Figure 7. Simplified use case diagram illustrating the interactions between the users and the various components of the e-BioInfra Gateway for computational neuroscience. For simplification, a link that is attached to a component means that it is attached to all actions within that package. The color code is the same as it is used in Figure 2.

4.7. Component interactions

Figure 7 illustrates the simplified use case of the SG and depicts the functionalities and services provided by the gateway components. User actions are expressed via the eBrowser and trigger interactions between other high-level components (i.e., PM, DTS, and eCAT) and lower-level components (i.e., gUSE and XNAT). User A is inside and User B is outside the AMC network. Details of these interactions are presented in the succeeding texts.

Upon successful authentication with the gateway, the user obtains access to the eBrowser portlet. New users need to configure an XNAT endpoint by providing their credentials. These configurations are collected by the eBrowser and sent to eCAT for validation and storage. After this configuration step, the following takes place when the user logs into the e-BioInfra Gateway.

At first, the user sees a list of her/his projects. To display this list, eBrowser sends a request to eCAT, which authenticates on behalf of the user to the XNAT and generates a list of all projects that are accessible by that particular user. Similarly, when the user selects a project, the eBrowser sends a request to eCAT, which queries meta-data on the XNAT to produce the list of all data entries in that project. Note that data and meta-data should be inserted into the XNAT prior to its retrieval from the gateway.

The user then selects data entities for processing and browses for available applications. The eBrowser retrieves and displays the list of applications that can be executed by the user, only showing applications that are compatible with the selected data type and format. The user selects an application, and the eBrowser displays configurations (application parameters) for that application. The user configures the application and starts a new data processing. The eBrowser collects the provided configuration and submits a processing request to the PM. The PM consults eCAT to find the details of the selected application, namely, the DCI to run it and the arguments that need to be

configured for its execution (e.g., input files and parameters). The PM creates a new processing entity in eCAT, from which the eBrowser can later retrieve and display to the user for browsing, management, and monitoring purposes.

The PM further instructs the DTS to move the required input data to the target DCI. The DTS contacts eCAT to determine if those data already have a replica on the target DCI. If no replica is available, the eCAT provides DTS with the XNAT endpoint configurations (including authentication token) and location where it can retrieve the input data. The DTS then uses this information to authenticate on behalf of the user to the XNAT and download the input data to the gateway server. Similarly, it retrieves user authentication tokens for the target DCI to upload input data. Finally, the DTS registers in eCAT the location of the file replica in the DCI and returns it to the PM.

After all data have been staged to the target DCI, the PM imports the application from gUSE via the ASM API, and configures it with the location of input data and user-specified parameters. Having everything in place, the PM starts the data processing by submitting the configured workflow to gUSE via the ASM API and updates the processing status in eCAT. The gUSE workflow interpreter parses the workflow, generates corresponding jobs, and submits them to the DCI-BRIDGE. This service then retrieves user-specific or community-specific authentication tokens for the target DCI to submit jobs on behalf of the user to the target DCI.

The PM periodically updates the information in eCAT based on the status reports from gUSE. The user browses, manages, and monitors the processing via the eBrowser. eBrowser contacts eCAT to obtain information about processing entities, including status. Different levels of details (views) are shown to the user and to the system administrator.

Typically, each processing consists of multiple data to be processed. When the processing of a data item is finished, the result is immediately stored in the XNAT via the DTS. Provenance data are associated to the results to identify the application that generated it and the input data. Thereby, the user can check results even before the entire processing is complete. The links to the results on the XNAT are displayed by eBrowser.

4.8. The e-BioInfra gateway for computational neuroscience

The new e-BioInfra Gateway for computational neuroscience is available on-line via <http://neuro.ebioscience.amc.nl>. The links to its source code, developer, and user documentations are also available on the SCientific gateway Based User Support portlet repository [21].

Currently, the following medical image processing applications are available to process diffusion tensor imaging (DTI) and structural magnetic resonance imaging data: (1) *Freesurfer*, implements segmentation of structural magnetic resonance imaging data with the Freesurfer toolbox [22]; (2) *DTI-preprocessing*, performs format conversion and quality control of DTI data [23]; and (3) *Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques for modelling crossing fibres (BEDPOSTX)*, performs local modeling of diffusion parameters with Software Library (FSL) BEDPOSTX [24].

5. USER FEEDBACK

The new gateway has been only recently released, in November 2013, for widespread usage at the AMC; therefore, it has not been possible so far to carry out a significant user study. Nevertheless, in this section, we describe our initial attempts to collect user feedback about the new gateway in various opportunities.

The new gateway has been first released for AMC users in July 2013. It has been thoroughly evaluated by two power users from the AMC Radiology department for a few months (July–September), during which extensive feedback was provided and the necessary improvements were implemented in the system.

After that, the new gateway was used in the ‘Advanced Cognitive Neurobiology & Clinical Neurophysiology’ course of the Biomedical Sciences Master program of the University of Amsterdam. After 1 h of introduction about e-Science and the gateway, the students used the gateway during another hour to perform high-level tasks such as ‘find the scan of your brain’, then

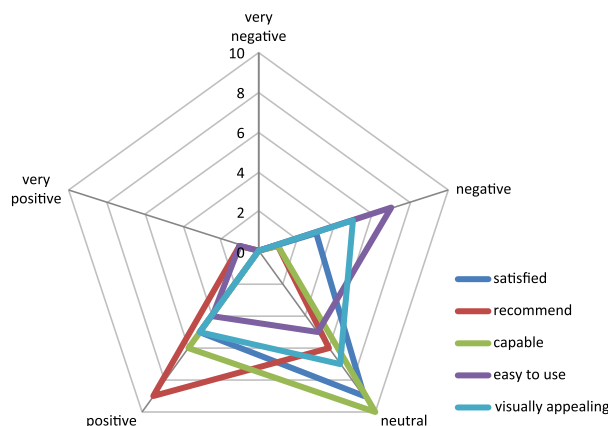


Figure 8. Summary of the feedback received from the students of the ‘Advanced Cognitive Neurobiology & Clinical Neurophysiology’ course after 1 h usage.

‘run Freesurfer on it’. During this course, 17 students used the gateway simultaneously through six user accounts. All students were able to successfully complete the data analysis tasks.

After the course, the students were asked to answer a questionnaire as the first external users of the system, and to give feedback about their experience with the new gateway. They answered the following questions using five multiple choices ranging from *very negative* to *very positive*:

- Overall, how *satisfied* or dissatisfied are you with the Neuroscience Gateway?
- How likely are you to *recommend* the Neuroscience Gateway to a friend or colleague?
- How *capable* is the Neuroscience Gateway in supporting your needs?
- How *easy to use* do you consider the Neuroscience Gateway?
- How *visually appealing* or unappealing do you consider the Neuroscience Gateway?

Figure 8 summarizes the responses of the students in a radar chart. Although these responses can only serve as a very initial assessment, they show no extremes. Note that, although most of the students found the gateway not easy to use, almost all of them indicated that they are likely to recommend it to others. We recall that the students, who were absolute beginners in the topic and the usage of e-Science environments, were able to complete the assignments successfully. This indicates that the gateway is easy to use for management of computational neuroscience data analysis and hides the complexities of underlying framework from the end-users. Additionally, on average, they were neutral about how satisfactory, capable, and visually appealing the gateway is.

Finally, in November 2013, the new gateway was officially released to the Brain Imaging Center of the AMC. During that event, the potential users pointed out that the interface was indeed looking better, but that hands-on experience with a real scientific task would be necessary for further feedback. Some of the users also pointed out the need to support other data sources, and to facilitate import of non-Digital Imaging and Communication in Medicine data into XNAT.

6. RELATED WORK

Design, development, and usage of SGs have gained interest and attention in the past few years. Several projects and initiatives have been started worldwide to develop SG frameworks and SG instances for diverse user communities [25]. For example, see the list of SGs on the websites of Extreme Science and Engineering Digital Environment [26], EGI [27], and the SCI-BUS [28] project. In particular, several neuroscience research communities have developed various SGs to integrate their medical imaging applications and data with access to computing and storage resources. This section presents some of the most recent SGs for the neuroscience research and how they related to our work.

The neuGRID for you (N4U) SG [29] provides user-friendly access to a large number of tools, algorithms, pipelines, visualization toolkits, data, and resources on various DCIs (grid, cloud, and

clusters) for medical imaging research. The goal is to exploit these tools towards the cure of neurodegenerative diseases, in particular Alzheimer's, psychiatric, and white matter disease. The N4U Persistency Service registers distributed data from project partners into the N4U Information Base, which are then treated as a single data source.

The CBRAIN portal provides transparent access to remote resources to manage, share, process, and visualize imaging data [30]. The CBRAIN platform links several brain imaging centers to high-performance computing (HPC) and cloud facilities across Canada and the world, both for data sharing and distributed processing. The data transfer and job submission details are transparently handled by the platform. Gee et al. [31] designed and implemented a data mining platform for neuroimaging data warehousing and processing that aim at brain recovery research. This platform integrates with CBRAIN for data processing and utilizes XNAT for data storage and sharing.

The Laboratory Of Neuro Imaging pipeline environment facilitates the integration of disparate data, tools, and services in complex neuroimaging data processing workflows. It supports neuroscientists with visual tools for data management and integration, and workflow development and execution on HPC platforms. It also updates the data provenance automatically during the processing [32].

The Virtual Imaging Platform portal [33] is a multi-modality medical image simulation platform that facilitates sharing of object models and medical image simulators. The models are described with semantic web ontologies and shared in a common repository. Virtual Imaging Platform portal enables users to run simulations implemented as MOTEUR workflows on the EGI computing resources and in-house clusters. Data are uploaded via the portal web interface using a dropbox-like approach; it is then stored on the EGI data resources and indexed in a central logical file catalog.

The Charité Grid portal [34] enables its users to run medical imaging applications on the German grid resources provided by the MediGrid and PneumoGrid projects. It also provides an interface to a picture archiving and communications system that contains anonymized medical images. The users are also provided with interfaces to upload data to the portal server. High-level data services upload data from the Picture Archiving and Communications System or the portal server to the grid computing resources for processing, and download the results to their desktop computer.

The Diagnostic Enhancement of Confidence by an International Distributed Environment (DECIDE) SG [35] provides high-level services for computer-aided neurological diseases diagnosis and research on the European Research and Education Networks, and EGI. It is based on the Catania SG framework [5] and utilizes a data engine that enables data transfer and sharing on grid storage resources.

The Neuroscience Gateway [36] is based on the Cyberinfrastructure for Phylogenetic Research (CIPRES) SG framework [37]. It enables the users to run parallel neuronal simulation tools on HPC platforms in the US cyberinfrastructure. It hides the technical details from the scientists for running jobs and managing data.

These SGs are usually designed and implemented based on the requirements of the specific research community that they support. Therefore, each one is unique in its own way. However, most of them display a few common characteristics that resemble our new gateway: data resources are directly connected to the gateway; a large variety of neuroimaging applications are available for the users; grid and clusters are used for high throughput computing. The major differences lie on the software platform used to implement the gateways, which varies from customized solutions to WfMS and SG frameworks. More information about the implementation would be necessary for a proper comparison of these systems; however, such information is normally not presented in the publications accessible to us, and, when they are, they may become obsolete very quickly.

7. DISCUSSION

The new gateway is significantly different from the previous one. Table I highlights the main differences concerning their main features. Although some of the features are similar, in many cases, they have been integrated. Moreover, their implementation is totally different in the two generations of e-BioInfra Gateway. The previous one was built based on the Spring framework, MOTEUR WfMS, and DIANE pilot job framework; it only supported the Dutch grid infrastructure,

Table I. Main differences between the previous and the new e-BioInfra Gateway.

Feature	Previous gateway	New gateway
Supported DCIs	Dutch grid	Dutch grid + in-house cluster
Supported external data servers	Absent	XNAT
Data transport	Semi-automatic and passive (GridSync)	Automatic and active
Search data based on meta-data	Absent	Present
Collaboration between researchers	Absent	Data shared within XNAT Projects
Provenance	Only data history	Full support (data, parameters, etc.)
Recovery from failures	External to the gateway	Integrated into the gateway
Communication between users and system administrators	External to the gateway	Integrated into the gateway
Workflow details and error messages	Exposed to the users	Hidden from the users

DCI, distributed computing infrastructures; XNAT, eXtensible Neuroimaging Archive Toolkit.

and it lacked facilities for data management, UI customization, or community support. In contrast, the new one is built based on the WS-PGRADE/gUSE SG framework, which itself is built on the Liferay portal framework. Liferay provides facilities for user management, community management, and community support (e.g., on-line forum). Moreover, it also facilitates the construction of customizable web-based UIs that are required to suit needs of each user (community) based on their profile, expertise, and roles. The WS-PGRADE/gUSE SG framework provides high-level generic services to manage workflows, enact them to various DCIs, and monitor their execution. These services allow for functional scalability and interoperability between various DCIs. Additionally, the WS-PGRADE/gUSE framework is an actively maintained, supported, and developed open-source project, which allows the development team of the e-BioInfra Gateway to concentrate on community-specific features, and makes the gateway maintenance more sustainable.

Currently, only XNAT is supported as data server. Several other data management platform alternatives meet the research requirements, although XNAT is of special interest due to its support for medical imaging. More importantly, it is already adopted by the AMC neuroscience research community. XNAT has been particularly designed for managing standard medical imaging data as the core of its functionalities. In addition, its archiving and integrating capabilities, data model flexibility, ease of use, and the highly active community of users/developers makes it a relevant asset. By connecting the XNAT to the e-BioInfra Gateway, the XNAT usage is also improved. Researchers are now able to perform compute-intensive data analysis on the data and receive the results in the same system. An alternative would be to develop XNAT pipelines to send processing jobs to external computing resources. Note however that the new e-BioInfra Gateway has been designed to support multiple and heterogeneous data servers, and it is not dependent on XNAT.

In this implementation, we chose to use an external data server, and keep the access control to this data server completely in the hands of the community administrators. This helped us build trust between the systems, which is a known critical factor to connect them to open infrastructures. Additional advantages of relying on external data servers for data management are flexibility, extensibility, data federation, and transfer of operational responsibilities to data owners. On the other hand, it is challenging because of issues such as connectivity, speed, and synchronization. For example, we experienced difficulties to connect to the XNAT server at the AMC because of the firewall policies. Also, the eCAT caches meta-data to reduce the frequency of XNAT queries; however, keeping the meta-data on XNAT and eCAT in sync also turned out to be challenging. Finally, if the external data server is discontinued or is off-line for any reason, the links from the eCAT to XNAT and much meta-data (e.g., provenance) present on the gateway become invalid.

We used WS-PGRADE/gUSE as SG framework, which in principle provides the workflow management and portal functionalities needed for our new gateway. After a learning phase, during which the concepts of the framework were better understood by the team, we observed that the usage model of the framework differs from our needs in some cases, which has led us to develop our own *processing manager* component. This has the goal of translating high-level 'data processing' commands into low-level data transports calls to the *data transport service*, and to

workflow execution calls to the gUSE ASM API. At first sight, this introduces small overhead, but at the same time, it provides sufficient isolation from aspects regarding this particular WfMS, and allows us to consider other WfMSs in the future. Moreover, this solution handles the data transfer and provenance collection properly, which would be more difficult to implement without this abstraction layer.

The development of eBrowser viewing portlets was also simplified by the decision to have all user interaction to take place using information available on the eCAT. This approach requires all software components to register all activity on the eCAT, but it decouples the viewer from all the other components accordingly. This reduces dependencies between the system components and simplifies its implementation and maintenance. Moreover, it makes eCAT a natural provenance data repository for the activity carried out at the gateway. The provenance is captured during the runtime, and the information that is relevant for the research community is stored in their data server as meta-data.

Among the motivations for the new gateway was the need to reduce the number of errors and also to handle them in a more elegant way. In the new gateway, the data-related errors are not observed by the end-users anymore because the gateway prevents them from processing incompatible data type and formats with its applications. Additionally, in case of an error, the submission is put on hold, and the administrators are notified about the error. This allows experts to inspect the error and act upon it (e.g., blacklist a faulty cluster on the grid and resume the submission) without involving the end-user unnecessarily. In the new gateway, the user is exposed to high-level user-friendly error messages only when there is some application or data-related problem that he/she can resolve.

8. CONCLUSIONS

In the new generation of the e-BioInfra Gateway, we tried to reduce the gap between users, data services, and DCIs. In contrast to our previous gateways, here, we aimed for a data-centric gateway in which everything is organized around 'data' and most importantly 'meta-data'.

Now users can use the gateway to browse their data and meta-data, which can be potentially stored on several data servers and described by rich meta-data, and to perform large scale data processing on them using DCIs. This can be performed without getting involved into low-level details of the infrastructure.

By making the gateway meta-data rich, the execution of applications is also streamlined. Because there are now more meta-data available about the input data and the applications, it is possible to assist the user in choosing the correct application. For example, if a data item is not compatible with a specific application, the gateway prevents the user from starting a processing for this combination. Moreover, applications are not isolated from each other any more. The output of one application is transferred to the XNAT with proper meta-data, which can be used to match this result to inputs of another application as a subsequent step in the data analysis pipeline. Although in some cases the steps can be linked at the workflow representation, in medical imaging it is still usual to visually check the results for quality control, which hampers full automation.

In the near future, the gateway will be disseminated in more training events, and become open to the whole neuroscience research community of the University of Amsterdam. This step will require inclusion of other data servers, for example, other XNAT instances or even other systems, as well as extending the eCAT with federated services for accessing (and/or querying) multiple data servers. Increasing number of users and data will possibly require further development of instruments for strong community support and communication. Moreover, semantic content annotation (ontologies), as well as adding knowledge and integrating it with existing data, could enable further automation of the data processing to reduce even more human intervention in the analysis of large quantities of biomedical data.

Finally, we kept bioinformatics researchers in the loop during the requirement analysis, design, and implementation of the gateway. The goal was to assure a design that is generic enough to support this new community with minimal additional effort. Although in this paper we focused on the computational neuroscience applications, the same concepts and software components are being used to develop an SG for protein docking.

ACKNOWLEDGEMENTS

We thank the colleagues who participated in the development, deployment, and testing of the new e-BioInfra Gateway for computational neuroscience: Mark Santcroos, Hurng-Chun Lee, Juan Luis Font, Paul F.C. Groot, and Gerbrand Spaans.

This work performed computations using resources of the Dutch e-Science Grid, which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO) and by Stichting SURF.

This work is financially supported by the COMMIT project ‘e-Biobanking with imaging for healthcare’ funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO); the SCI-BUS project funded by European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 28348; the ER-Flow project funded by FP7 INFRASTRUCTURES-2012-1 call under contract no. 312579; and the HPCN University of Amsterdam project ‘Computational Neuroscience Gateway’ funded by the University of Amsterdam.

REFERENCES

1. De Roure D, Jennings NR, Shadbolt NR. The semantic grid: past, present, and future. *Proceedings of the IEEE* 2005; **93**(3):669–681.
2. Hey T, Trefethen AE. Cyberinfrastructure for e-science. *Science* 2005; **308**(5723):817–821.
3. European Grid Infrastructure Science Gateway Virtual Team. *Science gateway primer*, 2012. (Available from: <https://documents.egi.eu/document/1463>) [Accessed on 1 December 2013].
4. Kacsuk P, Farkas Z, Kozlovsky M, Hermann G, Balasko A, Karoczkai K, Marton I. WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities. *Journal of Grid Computing* 2012; **10**(4):601–630.
5. The catania science gateway framework. (Available from: <http://www.catania-science-gateways.it>) [Accessed on 1 December 2013].
6. McLennan M, Kennell R. Hubzero: a platform for dissemination and collaboration in computational science and engineering. *Computing in Science Engineering* 2010; **12**(2):48–53.
7. Maddineni S, Kim J, El-Khamra Y, Jha S. Distributed application runtime environment (DARE): a standards-based middleware framework for science-gateways. *Journal of Grid Computing* 2012; **10**(4):647–664.
8. Shahand S, Santcroos M, Kampen AHC, Olabbarriaga S. A grid-enabled gateway for biomedical data analysis. *Journal of Grid Computing* 2012; **10**(4):725–742.
9. *The SURFsara website*. (Available from: <http://www.surfsara.nl>) [Accessed on 1 December 2013].
10. Peters BD, Machielsen MWJ, Hoen WP, Caan MWA, Malhotra AK, Szeszko PR, Duran M, Olabbarriaga SD, de Haan L. Polyunsaturated fatty acid concentration predicts myelin integrity in early-phase psychosis. *Schizophrenia Bulletin* 2013; **39**(4):830–838.
11. van Wingen GA, Geuze E, Caan MWA, Kozicz T, Olabbarriaga SD, Denys D, Vermetten E, Fernández G. Persistent and reversible consequences of combat stress on the mesofrontal circuit and cognition. *Proceedings of the National Academy of Sciences* 2012; **109**(38):15508–15513.
12. Rienstra A, Groot PFC, Spaan PEJ, Majoie CBML, Nederveen AJ, Walstra GJM, de Jonghe JFM, van Gool WA, Olabbarriaga SD, Korkhov VV, Schmand B. Symptom validity testing in memory clinics: hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology* 2013; **35**(1):59–70.
13. de Kwaasteniet B, Ruhe E, Caan M, Rive M, Olabbarriaga S, Groefsema M, Heesink L, van Wingen G, Denys D. Relation between structural and functional connectivity in major depressive disorder. *Biological Psychiatry* 2013; **74**(1):40–47.
14. *The BIC (Brain Imaging Center) at the AMC (Academic Medical Center) website*. (Available from: <http://www.lebic-amc.nl>) [Accessed on 1 December 2013].
15. Glatard T, Montagnat J, Lingrand D, Pennec X. Flexible and efficient workflow deployment of data-intensive applications on grids with MOTEUR. *International Journal of High Performance Computing Applications* 2008; **22**(3):347–360.
16. Moscicki JT, Lamanna M, Bubak M, Sloot PMA. Processing moldable tasks on the grid: late job binding with lightweight user-level overlay. *Future Generation Computer Systems* 2011; **27**(6):725–736.
17. Madougou S, Shahand S, Santcroos M, van Schaik B, Benabdalkader A, van Kampen A, Olabbarriaga S. Characterizing workflow-based activity on a production e-infrastructure using provenance data. *Future Generation Computer Systems* 2013; **29**(8):1931–1942.
18. Digital Imaging and Communications in Medicine. (Available from: <http://medical.nema.org>) [Accessed on 1 December 2013].
19. Shahand S, Caan MW, van Kampen AH, Olabbarriaga SD. Integrated support for neuroscience research: from study design to publication. *Studies in Health Technology and Informatics* 2012; **175**:195–204.

20. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The extensible neuroimaging archive toolkit. *Neuroinformatics* 2007; **5**:11–33.
21. Amc neuroscience gateway portlet on the sci-bus portlet repository. (Available from: <https://scibus-repo.cpc.wmin.ac.uk/scibus-repo/public/details-view.xhtml?portletID=4353>) [Accessed on 1 December 2013].
22. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. automatically parcellating the human cerebral cortex. *Cerebral Cortex* 2004; **14**(1):11–22.
23. DTI Preprocessing on the e-BioinfraGateway. (Available from: <http://www.bioinformaticslaboratory.nl/twiki/bin/view/EBioScience/PredtiUserDoc>) [Accessed on 1 December 2013].
24. FMRIB's Diffusion Toolbox - BEDPOSTX. (Available from: http://fsl.fmrib.ox.ac.uk/fsl/fsl4.0/fdt/fdt_bedpostx.html) [Accessed on 1 December 2013].
25. Kiss T. Science gateways for the broader take-up of distributed computing infrastructures. *Journal of Grid Computing* 2012; **10**:599–600.
26. The XSEDE (Extreme Science and Engineering Digital Environment) website. (Available from: <http://www.xsede.org>) [Accessed on 1 December 2013].
27. EGI (European Grid Infrastructure) Science Gateways. (Available from: <http://go.egi.eu/sciencegateways>) [Accessed on 1 December 2013].
28. The SCI-BUS (SCientific gateway Based User Support) Project website. (Available from: <http://www.sci-bus.eu>) [Accessed on 1 December 2013].
29. The N4U (neuGRID for you) Project website. (Available from: <http://neugrid4you.eu>) [Accessed on 1 December 2013].
30. Frisoni GB, Redolfi A, Manset D, Rousseau ME, Toga A, Evans AC. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. *Nature Reviews Neurology* 2011; **7**(8):429–438.
31. Gee T, Kenny S, Price CJ, Seghier ML, Small SL, Leff AP, Pacurar A, Strother SC. Data warehousing methods and processing infrastructure for brain recovery research. *Archives Italiennes De Biologie* 2010; **148**(3):207–217.
32. Dinov I, Lozev K, Petrosyan P, Liu Z, Eggert P, Pierce J, Zamanyan A, Chakrapani S, Van Horn J, Parker DS, Magsipoc R, Leung K, Gutman B, Woods R, Toga A. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS ONE* 2010; **5**(9):e13070.
33. Glatard T, Lartizien C, Gibaud B, Ferreira da Silva R, Forestier G, Cervenansky F, Alessandrini M, Benoit-Cattin H, Bernard O, Camarasu-Pop S, Cerezo N, Clarysse P, Gaignard A, Hugonnard P, Liebgott H, Marache S, Marion A, Montagnat J, Tabary J, Friboulet D. A virtual imaging platform for multi-modality medical image simulation. *IEEE Transactions on Medical Imaging* 2013; **32**(1):110–118.
34. Wu J, Siewert R, Hoheisel A, Falkner J, Strauß O, Berberovic D, Krefting D. The charité grid portal: user-friendly and secure access to grid-based resources and services. *Journal of Grid Computing* 2012; **10**:709–724.
35. Ardizzone V, Barbera R, Calanducci A, Fargetta M, Ingrà E, Porro I, La Rocca G, Monforte S, Ricceri R, Rotonondo R, Scardaci D, Schenone A. The decide science gateway. *Journal of Grid Computing* 2012; **10**(4):689–707.
36. Sivagnanam S, Majumdar A, Yoshimoto K, Astakhov V, Bandrowski A, Martone MA, Carnevale NT. Introducing the neuroscience gateway. *Proceedings of the 5th International Workshop on Science Gateways*, Zurich, Switzerland, June 2013.
37. Miller MA, Pfeiffer W, Schwartz T. The cipres science gateway: a community resource for phylogenetic analyses. In *Proceedings of the 2011 Teragrid Conference: Extreme Digital Discovery*, TG '11. ACM: New York, NY, USA, 2011; 41:1–41:8.