The final ACGT portal, and online training modules development and evaluation

Project Number: FP6-2005-IST-026996

Deliverable id: D 14.8

Deliverable name: The final ACGT portal, and online training modules development and evaluation

Date: 31 July, 2010

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery
Document id:	D 14.8
Document name:	**The final ACGT portal, and online training modules development and evaluation**
Document type (PU, INT, RE)	RE[1]
Version:	2.0
Date:	31.07.2010
Editor: Organisation: Address:	Alexandru Sofronia SIVECO 73-81 Bucharest-Ploiesti, Bucharest, Romania

Document type PU = public, INT = internal, RE = restricted

ABSTRACT:

The present deliverable is the final report on the ACGT Portal following *D14.2 Visual prototype and report of the ACGT portal*. The report is a merge of two previously separated deliverables: *D14.6 First Report on ACGT Portal usage, and the development and evaluation of online training modules* (released in November 2008 for internal review, but not submitted for EC review) and *D14.8 Second Report on Second Report on ACGT Portal usage, online training modules development and evaluation.*

The ACGT Portal is currently in a stable phase (third - and final - prototype), after the integration of different heterogeneous services developed in the ACGT Portal, while the development of training materials has also reached a stable phase – training materials have been developed and used in training activities (see D14.4). The evaluation of the ACGT Portal which was carried out in the training and evaluation activities is covered in deliverable D13.4.

KEYWORD LIST: Web Portal, Grid Portal, Data Mining Portal, Knowledge Discovery Portal, Online Training Modules

[1] Deliverable is RE, but public information contained in this deliverable is available in a open-community Wiki – ACGT Handbook - http://handbook.eu-acgt.org/

MODIFICATION CONTROL			
Version	Date	Status	Author
0.1	10.09.08	Draft (as D14.6)	Radu Gramatovici
0.9	29.11.08	Draft (as D14.6)	Radu Gramatovici
1.0	30.11.08	Release for internal review (as D14.6)	Radu Gramatovici
1.1	30.06.09	Draft (as D14.8)	Alexandru Sofronia
1.2	01.10.09	Draft (as D14.8)	Alexandru Sofronia
1.5	31.01.10	Draft (as D14.8)	Alexandru Sofronia
2.0	31.07.10	Release (as D14.8)	Alexandru Sofronia

List of Contributors

- Radu Gramatovici, SIVECO Romania
- Alexandru Sofronia, SIVECO Romania
- Sorin Portase, SIVECO Romania
- Andreea Tabacariu, SIVECO Romania
- Otto Zelch, SIVECO Romania
- Andreas Persidis, Biovista

Contents

Table of Figures

Executive Summary

ACGT is an Integrated Project (IP) funded in the 6th Framework Program of the European Commission under the Action Line "*Integrated biomedical information for better health*". The high level objective of the Action Line is the development of methods and systems for improved medical knowledge discovery and understanding through integration of biomedical information (e.g. using modelling, visualization, data mining and grid technologies).

In the architecture of the ACGT Grid infrastructure, the ACGT portal acts as the main gate to the knowledge-discovery tools developed in the ACGT portal. Knowledge Discovery can be described as deriving knowledge from input data. The ACGT Environment integrates different knowledge discovery tools and services that can be used from a single user interface: the ACGT portal. Structured on several layers of access rights, the ACGT Portal offers end-users and service-providers[2] access to different private areas. A complete overview of the main areas of the ACGT Portal is available in the ACGT Handbook wiki: http://handbook.eu-acgt.org/HB:ACGT_Data_Mining_Tools

This deliverable is the second report on the ACGT Portal. The first report on the ACGT Portal was presented in the deliverable *D14.2 Visual prototype and report of the ACGT portal*. The report is a merge[3] of two previously separate deliverables:

- *D14.6 First Report on ACGT Portal usage, and the development and evaluation of online training modules* (released in November 2008 for internal review, but not submitted for EC review)

- *D14.8 Second Report on Second Report on ACGT Portal usage, online training modules development and evaluation.*

The ACGT Portal is currently in a stable phase, after the integration of different heterogeneous services developed in the ACGT Portal, while the development of training materials has also reached a stable phase – training materials have been developed and used in training activities[4]. The evaluation of the ACGT Portal which was carried out in the training and evaluation activities is covered in deliverable *D13.4 Final Evaluation Report*.

This deliverable is structured in two parts:

Part 1 of this deliverable covers the ACGT Portal development methodologies - spiral model, scenario-based approach, service integration issues - in chapter **2** and online training modules and materials in chapter **3**.

Part 2 of this deliverable covers the three ACGT Portal prototypes and a comprehensive list of front-end and back-end portlets in chapter **4**, ACGT Portal homepage design changes in prototypes in chapter **5**, ACGT Registration and login usability changes in prototype 3 in chapter **6**, front-end portlets and changes in the prototypes in chapter **7** and back-end ACGT portlets in chapter **8**. Some browser and network configuration settings for end-users are discussed in chapter **9**.

[2] For further reference on the ACGT roles: http://handbook.eu-acgt.org/HB:ACGT_Roles

[3] The deliverables were merged in DOW IP4

[4] For further reference on training activities see *D14.4 Report on training of end-users and service-providers on ACGT Technologies & Methodologies*

PART 1

ACGT portal and training modules development

1 Short description

The **ACGT Portal** is the interface of the ACGT environment, which attempts to provide an integrated easy-to-use and up-to-date gateway to the ACGT tools and services.

The main roles of the ACGT Portal are:

- It provides a unique access point for the ACGT Grid
- It provides a customizable client for ACGT services
- It integrates the usage of ACGT internal and external services
- It provides the main ACGT channel for training

In many respects, the ACGT Portal gives access to the services of the Business Process Layer and it **is the main point of entry** in the ACGT environment. The Knowledge Discovery and Data Mining Tools developed in the ACGT project are grouped and accessible in an integrated environment through the ACGT Portal.

The ACGT Portal provides a Certificate-based Single Sign on environment, allowing users to access all the services and data mining tools a unique registration process and a unique login process.

A *single registration process* is necessary to grant access to the ACGT environment. Because genuine personal data are flowing in the system, ACGT has a stronger authentication and authorization policy than many online services. The real identity of the users needs to be validated before access to the system is granted. After the registration, an *ACGT Passport* is issued and, just like an ordinary passport, users will be asked for it when crossing the border into the ACGT environment. However, to access the ACGT environment from a different computer than the original computer used to store the passport, users must: EITHER copy the *ACGT Passport* from the original computer to the other one or to a USB stick for convenience OR request temporary *ACGT Visas* – a temporary username and password pair which can be used for 7 days. These ACGT Visas can be used in a simplified login process (as in every-day portals).

After logging in to the ACGT Portal, all the private areas of the ACGT environment can be accessed. Some areas (usually those in a transitory phase, require group membership before allowing access.

The mapping between **what can be done** in the ACGT environment and **where it can be done** (the ACGT components involved) is presented in the diagram. This overview of the main (private) areas accessible in the ACGT Portal is also available as a clickable map on the main page of the portal ("*My ACGT*"), after logging in:

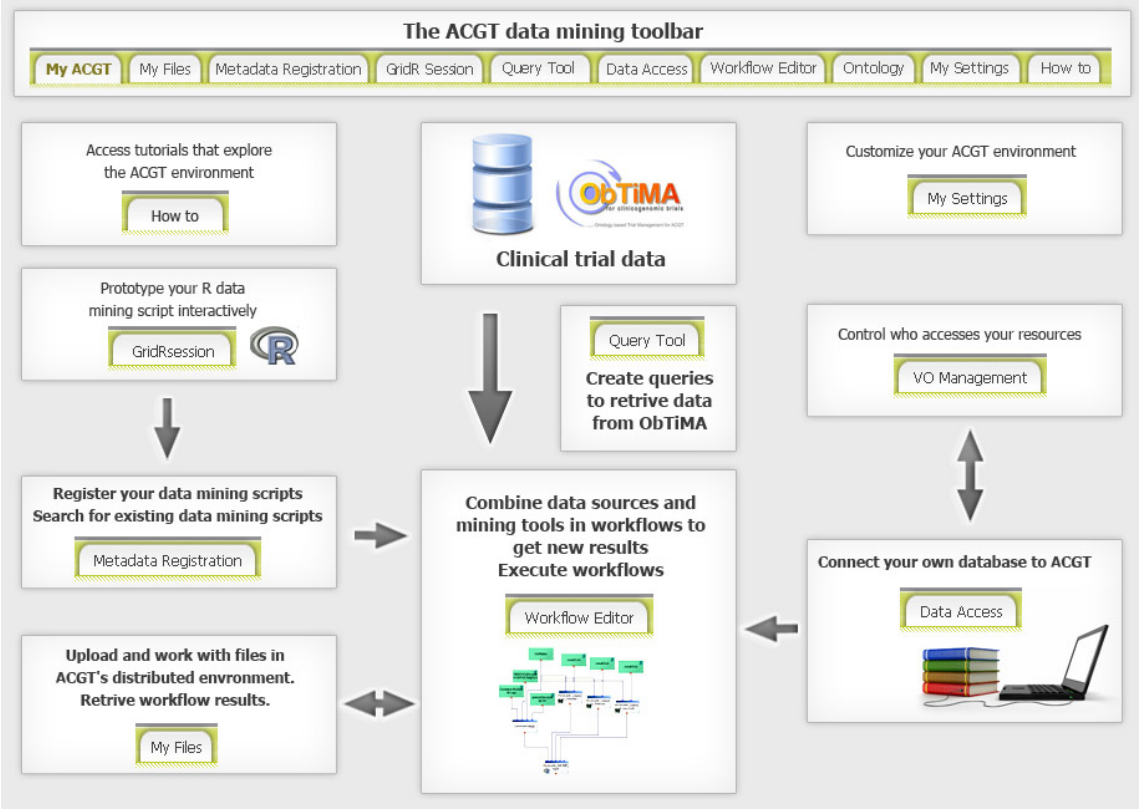

Figure 1 – ACGT Portal Overview

The ACGT Portal is built on the Gridsphere portal framework. GridSphere is an open-source JSR-168 compliant portal framework that is ready to run with a suite of tutorial and example modular web components, called portlets. A portlet is a module of a dynamic web page created using Java. For creating dynamic web pages, JSP (Java Servlet Pages) was developed, which combines HTML with Java - a JSP web server to run such applications is needed. As only some parts of the dynamic page actually needs to be modified while doing a set of actions, web page content was redesigned as a set of integrated dynamic JSP modules and static HTML content.

JSR (Java Portlet Specification) 168 is one of the main standards in portal development. **JSR 168** is a standard API for integrating portlets in portlet containers. JSR 168 defines a contract between the portlet container and portlets and provides a convenient programming model for portlet developers. The Java Portlet Specification V1.0 was developed under the *Java Community Process* (JCP) (having as members experts from leading industry companies) as JSR 168, and released in October 2003. The second version of the standard, Java Portlet Specification V2.0 or **JSR 286** was released in July 2008 and is in the process of adoption. JSR 286 is the Java Portlet specification v2.0 as developed under the JCP and created in alignment with the updated version 2.0 of WSRP. JSR 286 was developed to improve on the short-comings on JSR-168. Some of its major features include: Inter-Portlet Communication through events and public render parameters, serving dynamically generated resources directly through portlets, serving AJAX or JSON data directly through portlets, introduction of portlet filters and listeners.

The Gridsphere portal framework provides a standard based platform for the easy development of portlets. Portlets are defined by a standard API and provide a model for developing new portal components that can be shared and exchanged by various portlet containers. Gridsphere provides a portlet container, a collection of core portlets and an

advanced user interface library that makes developing new portlets easier for application developers.

Currently, three versions of the ACGT Portal are maintained:

- **Stable version**, which contains portlets providing access to services that are in a stable development state. This portal is used for demonstrations and user evaluations. This portal can be accessed both using HTTP and HTTPS protocols:

  http://acgt.siveco.ro/ and https://acgt.siveco.ro/

- **Stable version (mirror 1)**, used as a backup in demonstration and user evaluations. The address of this portal is:

  http://epimetheus.ics.forth.gr:8080/

- **Development version**, which contains portlets under development or portlets for services that are still under development. This portal is used for internal consortium-testing only. The address of this portal is:

  http://acgt.siveco.ro:8080/acgt/portal/

2 Development methodology

The keywords for the development of the ACGT Portal are: *spiral model* and *scenario-driven*.

2.1 ACGT Portal prototype development

The development of the ACGT Portal used a spiral model, based on several prototypes and with 1 year long iterations. Prototypes are initial versions of the system which are used for experimentation. These prototypes are valuable for requirements elicitation as users can experiment and discover what the real needs are, and more importantly, the spiral-prototype development model ensures in-depth requirements analysis which reveals inconsistencies and omissions.

The ACGT Portal is on the top layer of the ACGT architecture. Even if the development of the ACGT Portal in WP14 is independent from the rest of the ACGT system, the ACGT portal is highly influenced by the evolution of the rest of the components. For further details on the integration of ACGT Portal in the general ACGT architecture see *D3.1 – The ACGT Initial Architecture*.

The ACGT portal was not developed in one step, based only on the initial analysis. Instead, several prototypes of the portal were developed. At each stage more complexity was added to the portal both in terms of functionalities and technical integration. Gradual implementation of a portal keeps the complexity of the portal in reasonable limits and allows users to get familiar with a portal while it is growing in complexity.

At the same time, gradual implementation of the technical features of the portal into the grid system allowed early testing of the ACGT tools and services, even before their complete integration in the ACGT system. At the moment of the implementation in the portal prototype, some functionalities belonging to the grid layer may have not been in a stable and mature state and for this reason some of the portal features in earlier prototypes have been implemented in a transitory form at the portal level, rather than at the grid level – serving as “placeholders” for the final grid-level implementations.

As rapid development of the prototype is essential in the requirements elicitation process, the first ACGT portal prototype was developed early in the project and covered essential portal requirements (access, role-management, credential management, layout-configuration for user customization of the portal, etc). Its goal was to establish feasibility and usefulness before higher development costs are incurred. Its secondary goal was to establish a stable “look and feel” for the user interface. The first ACGT portal prototype is covered in-depth in deliverable D14.2.

The second and third ACGT portal prototypes integrated and provided different types of services, depending on their availability in the ACGT environment. A secondary goal of each prototype was to improve the usability of the portal, following evaluation results and user-feedback of the previous prototype.

2.2 Scenario-driven approach

In developing the ACGT portal prototype, several user scenarios have been taken into account. These scenarios have been established and refined in other WPs (mainly WP13).

Scenarios are stories, which explain how a system might be used. Telling stories about systems helps to ensure that project stakeholders share a sufficiently wide view to avoid missing vital aspects of problems. Scenarios vary from brief stories to richly structured analysis, but are almost always based on the idea of a sequence of actions carried out by intelligent agents.

Using a scenario-driven approach for the development of the ACGT portal we avoid a massive development of the portal in a wrong direction, adding new features in ACGT portal prototypes only when specific user scenarios required them.

The development of the second ACGT Portal prototype was driven by the following scenarios: *pseudo-TOP scenario* (aimed at testing a real current-day clinical trial with all the complexity of database), *data mining of public data* (creation of GridR-based service by an independent bioinformatician), *data mining of real data* (creation of GridR-based service by a bioinformatician attached to a clinical trial) and *MCMP pseudo-multicentric-multiplatform trial* (Affy/Illumina comparison - testing a study with large compute/storage requirements).

2.3 Service integration

From the service integration point-of-view, the ACGT Portal represents the Business Process Layer, in which services are primarily described from the user/usage point of view. Service registry is a component responsible for storing information about services available in ACGT environment. It provides enough information for enabling the execution of at least the following operations: registering, unregistering, updating, search and discovery, availability, access and monitoring.

From the integration point of view, the leading idea is that as many service tools as possible should be defined and implemented at grid level. The ultimate goal is to have as many services as possible fully integrated in the ACGT Grid. However, in a spiral model of development not all the useful/necessary services will be fully implemented earlier in the development phase in order to achieve some scenario-driven goals. Therefore, in some cases, some of these tools had to be defined and implemented directly in the ACGT Portal corresponding prototype.

More details on service integration in the ACGT Portal can be found in the deliverable *D14.5bis Methodology for ACGT service integration in the ACGT Portal on the Business Process Layer*.

Three different types of services can be accommodated in the ACGT Portal, reflecting different stages of the integration of these services in the ACGT system. These three different types were outlined in *D14.2 Visual prototype and report of the ACGT portal*:

- **ACGT integrated services** - services accessible through the ACGT Portal, that are fully compliant with the ACGT standards and quality requirements.

- **ACGT external services** - services provided by ACGT partners that are in a transitory phase which does not fulfill all of the requirements (technical, regarding security, etc) to be fully integrated in the ACGT infrastructure. For these services all

the tools will be available, but some of them might be defined in the portal. Once these services are fully integrated, they become ACGT integrated services.

- **Third-party services** - services registered by providers that do not belong to the ACGT Consortium. These services will not interfere with the core of the ACGT infrastructure, services and data. These services will be published under the responsibility of each provider and used under the responsibility of each ACGT user. Third-party services may become in time ACGT external or even integrated services, after the fulfillment of all necessary conditions.

Earlier ACGT Portal prototypes had most of the ACGT services as external services, providing different degrees of compatibility with the ACGT architecture. Even the latest ACGT Portal prototype has a mix of internal and external services, as some of these services will remain in a transitory form even at the end of the project.

3 Online training modules and materials

3.1 Embedded support and help documents

As stated in D14.3 *"Demonstration and Report of training modules"*, Part 1 "*Principles of online training modules*": when delivering the online help for using an IT application, users are not willing to frequently switch to a Help section or to frequently open an user manual. Instead, what they really need is to find fast and simple answers to the ordinary problems they encounter **when** performing specific actions.

Two of the most critical usability problems (Users not knowing how to begin interacting with the application and users misunderstanding the meaning and implications of pushing certain buttons) can be solved by embedding the help into the user interface, adding helpful information about the task at hand, right on the page, sometimes in the form of a patch on the screen that reveals the helpful text on mouse-over.

Embedded support should explain for the system itself (and each new area a user discovers):

- *How to begin interacting*: On the first screen that is displayed when opening the system or a new area for the first time, the user must be told what they can do and how to proceed.

- *What the buttons mean*: Instead of a one-word definition of each icon, which may be ambiguous for some users, tooltips are to be used to provide meaningful explanations of what each button does if pressed.

However, according to most usability experts embedded help systems should not be the only way to ease the adoption and use of a system. Most usability experts believe that if the user interfaces were properly designed, no help would be necessary. That level of perfection is seldom achieved in practice for complex systems which allow complex tasks to be performed. Therefore, necessary help information should be provided, without forcing the user to search for it.

Some simple design principles must be used throughout the development phase to ensure a maximum ease of adoption for the system:

- Make the information task-oriented and highly structured, because the purpose of online help is to get the user back on task as fast as possible

- Separate information into distinct information types and include only one information type in each online help topic. Information types can be by content type (task, concept, reference) or user models (beginner, intermediate, advanced).

- Follow minimalist principles in designing online help topics. Many online help designers misconstrue minimalism to simply mean brevity; however, minimalist design hinges on being able to make good decisions on what to do, say, or show, and on what not to include.

- Keep the length of topics short. However, there is no infallible length prescription: online help topics need to be "as long as necessary and as short as possible".

Other usability problems, such as users not knowing what else can be done with the system or having difficulties in finding information, can only be addressed through a better organization of and better access to the information.

For the ACGT Portal, since it contains a mix of portlets developed from-scratch for the ACGT Project and preexisting or third-party services (sometimes along with their own external user-interfaces), a trade-off must be reached in applying the above principles.

Some information, usually shorter in length information should be provided as embedded help, while the rest of the information is provided as large online help documentation and task or scenario-oriented tutorials.

3.2 ACGT Handbook

The ACGT project, due to the heterogeneous nature of the services and expertise and the involvement of a large number of partners, requires a larger and more flexible repository of information than traditional *User Manuals* or *Online Help* usually provide[5].

A wiki-based approach was used to provide consortium (and later user-community) reviewed content. The **ACGT Handbook** – http://handbook.eu-acgt.org/ - was started as a collaborative wiki, hosted by HealthGrid and edited by 5 members of the Technical Management Committee (SIB, FHG, PSNC, FORTH and chief-editor SIVECO).

Its main goals were to provide on-line help information, online and offline training materials supporting training sessions. Its secondary goals were to provide feedback over embedded help and user-interface issues which can be used for enhancements in later prototypes.

The ACGT Handbook is a *MediaWiki* v1.15.1 installation with *Parser Functions* and *PDFBook* extensions. The latter extension allows any Handbook page or set of pages to be converted and downloaded as a PDF file. Such files were used as handouts during training sessions.

The ACGT Handbook currently has **54** main content pages (**329** total number of pages), **155** images and has been visited **20010** times. **29** users are registered members of the community and have made **1265** page edits since the Handbook has been set up.

The ACGT Handbook is divided into 4 main areas:

- **ACGT Handbook for Knowledge Discovery and Data Mining** (namespace *HB:*) - pages dedicated to the ACGT environment accessible from the ACGT Portal – e.g. *ACGT Data Mining Overview, Getting started, ACGT Passport, Importing data, Storing data, Retrieving data, Browsing/registration of services, Browsing/registration of data types, Setting up workflows, Analyzing data with R, Working with the ACGT Ontology*

[5] Previously, a draft "*User Manual*" was initiated after Technical Management Chats and edited by several TMC members. This user manual has proved its usefulness by providing top-down skeleton pages and initial content for the *ACGT Handbook* initiative

- **ACGT Handbook for Trial Management** (namespace *HB:*) – pages dedicated to ObTiMA – e.g. *Registration, Login, Handbook for ObTiMA Administrator, Handbook for Trial Chairman, Handbook for Clinicians*

- **ACGT Programmer's Manual** (namespace *PM:*) - contains information for programmers (developers) who want to integrate new functionalities into the ACGT platform, in general any 3rd-party contributors – e.g. *Integrating R scripts, Developing a new service, Web Service Wizard, Developing a new Proxy Service, ACGT Test bed Security Guide, ACGT Bug Reporting - General Guidelines*

- **ACGT Examples** (namespace *EX:*) – which covers examples of how to solve real-life problems in clinico-genomic research using the ACGT platform, scenarios using the ACGT Environment and ACGT Tutorials – e.g. *ACGT End to End Scenario: From Trial Management to Data Mining, Integrated demonstrators* and 8 tutorials.

All the pages in the ACGT Handbook were updated and split between:

- *User Interface pages* - that are continuously updated to accommodate changes in the User Interface - e.g. *HB:Workflow_Editor*

- *task-oriented pages* - that tend to have a longer lifespan before requiring major updates - e.g. *HB:Setting_up_workflows*

A complete overview of the ACGT environment covers both UI and task-oriented pages: http://handbook.eu-acgt.org/HB:ACGT_Data_Mining_Tools

The ACGT Handbook wiki is an open-community, edits are allowed after an automated registration process (no anonymous edits). During the ACGT Competition, editorial rights were restricted to wiki Administrators only (SIVECO, HealthGrid, SIB) to prevent accidental or malevolent changes in online documentation (e.g. bot-spam), especially in the Programmer's Manual.

The ACGT Handbook is used as embedded help in the ACGT Portal as each corresponding portlet has a direct link (a blue question-mark icon) to the corresponding page in the Handbook. This allows users to quickly find the information related to the area of the ACGT environment they are visiting and also to find out what else can be done in ACGT – addressing the two usability problems which cannot be resolved by embedded help alone.

Moreover, the use of a wiki as a central information repository allows a large number of key-words (e.g. ACGT Passport) to have their own dedicated page which is available as a link providing further information on every page that mentions the key-word. Along with the previous types of help support which provide breadth-first access to information, the wiki also provides a depth-first way of finding information.

3.3 Video and multimedia tutorials

A third category of support materials and the most related to the notion of *online training* are video and multimedia tutorials. Each tutorial can provide task or scenario-oriented information in a more compelling way than any other form of presenting information.

Two types of tutorials have been deemed necessary for ACGT:

- **Public tutorials** - explaining what ACGT stands for, why should one use the ACGT environment, and what one must do to access the services provided by ACGT – e.g. registration, login. A most notable tutorial is the ACGT Educational Video which is covered separately in D14.7.

- **Private tutorials** - explaining how to use different ACGT tools and services and therefore targeting each group of users potentially interested in the corresponding tools and services.

Due to the prototype nature of the ACGT environment, all the private tutorials *are also openly available* from the ACGT Portal home page (since the second prototype), in an effort to convince potential users of the benefits of using different areas of the ACGT environment.

Previous sample tutorials were documented in *D14.3 Demonstration and Report of training modules:*

- Sample task-oriented private tutorials:
    - ACGT-WRT-SIV-v1 – *How to register a workflow?*
    - ACGT-OV-BIOVISTA-v1 - *Ontology Viewer*
- Sample tutorial covering external services:
    - ACGT-MAG-UMA-v1 *Composing Complex Workflows with Magallanes*
    - ACGT-jOrca-UMA-v1 *Creating an Aminoacid Sequence with jOrca*
    - ACGT-jOrca-UMA-v2 *Launching Services with jOrca*
    - ACGT-jOrca-UMA-v3 *Launching GRID Services with jOrca*
- Scenario-oriented private tutorials
    - ACGT-SCEN-SIV-v1 - *Data mining tools tutorial*[6]

Beyond these sample tutorials and in an effort to cover the most of the ACGT environment, a complete end-to-end scenario "*From Trial Management to Data Mining*" was developed by SIB, tested by HealthGrid and SIVECO and thoroughly documented in the ACGT Handbook by SIVECO (http://handbook.eu-acgt.org/EX:ACGT_End_to_End_Scenario) as step-by-step procedures which cover most uses of the ACGT platform.

The scenario was covered in 6 training modules (task-oriented tutorials) that were developed as SWF Flash movies using Adobe Captivate: *First time user: Register and get your ACGT Passport; Login with your Passport / Visa; Querying clinical data; Register and access external databases; Register and use GridR services; Create and use workflows*

A list of these task-oriented tutorials along with a short description is presented:

- **ACGT Registration** - The *ACGT Knowledge Discovery/Data Mining Tools* and additional resources are grouped and accessible in an integrated environment through the ACGT Portal. A single registration process is necessary to grant access to the ACGT environment. This tutorial shows the steps needed to get your ACGT Passport. Further information can be found here: HB:ACGT_Registration.

[6] This tutorial was used in Vienna, January 2009. As most parts of the tutorial are obsolete in the third prototype, and the scenario became more complex, it was replaced by a series of consecutive task-oriented tutorials rather than a single bulk tutorial.

- **ACGT Login** - The ACGT Knowledge Discovery/Data Mining Tools and additional resources are grouped and accessible in an integrated environment through the ACGT Portal. This tutorial shows how to login using your ACGT Passport and how to create short-time Visas to login easier the next time or to login from other computers. Further information can be found here: HB:Logging_in_and_out.

- **Querying clinical data** - In the ACGT environment, data from different trials can be accessed "as if" it was stored in a single integrated repository. The ACGT Query Tool can be used to retrieve data from clinical registered trials by defining queries and testing them. You can register these queries in ACGT and use them to get clinical data in workflows.

- **Register and access external databases** - In addition to the databases that are connected statically and with a fixed ontology mapping, users of the ACGT environment can also connect SQL databases dynamically to the data mining tools. The Data Access area allows users to add new databases and to execute queries on previously added dynamic databases. These databases can then be used in more complex workflows as data pools.

- **Register and use GridR services** - You can add analytical components that use R programming language (R scripts) to ACGT in order to analyze data. This tutorial shows how to add GridR scripts as ACGT available services that can be used in workflows. The example in the tutorial can be used in workflows for the analysis of the association between a patient's treatment with a specific drug and tumor relapse.

- **Create and use workflows** - Workflows can be designed to use and reuse different data pools and knowledge discovery and analytical services for more complex research goals. In this example, a GridR service uses both clinical trial data (3 mediator queries created in the Query Tool) and data from an external database which is queried using SPARQL to plot the association between patient's treatment and tumor relapse.

More details on the development and use of the training materials can be found in the deliverable *D14.3 Demonstration and Report of training modules* and *D14.4 Report on training of end-users and service providers on ACGT Technologies & Methodologies.*

3.4 Accessing help and tutorial in the ACGT Portal

Help and tutorials in the ACGT Portal are available for both for registered and unregistered users.

For *unregistered users*, the main page of the portal shows mostly public information:

- A short description of ACGT with links to clinical trials and grid computing key-notes.

- A step-by-step clickable procedure for unregistered users, with additional (*More info*) links to the ACGT Handbook corresponding pages:

    1. *Get an ACGT Passport* (registration)

    2. *Get a Visa and Login* (login)

3. *Enjoy the ACGT environment* (overview)

- A short-list of available ACGT Tutorials and resources
- Other help documents in PDF or SWF format

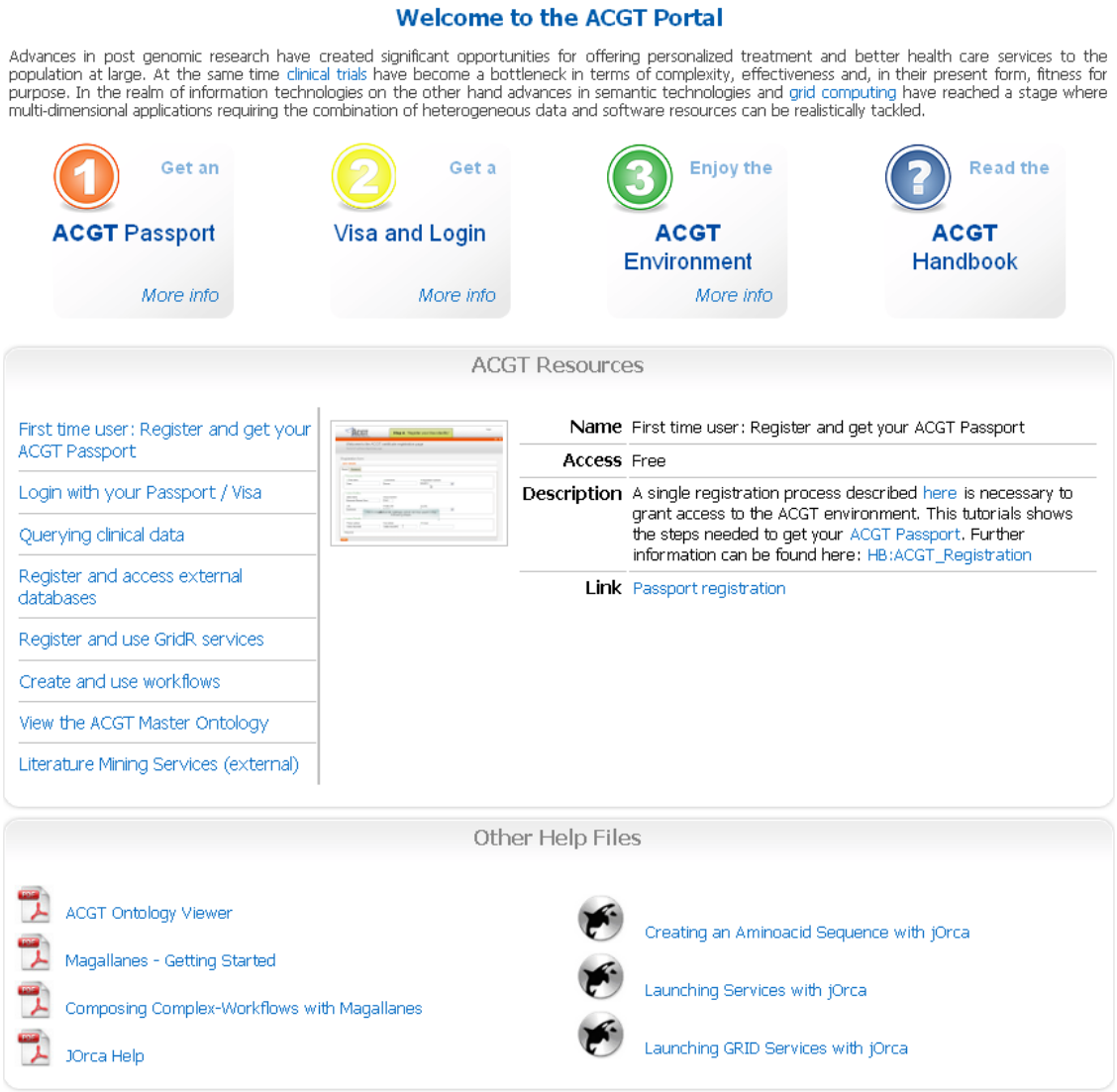

Figure 2 – ACGT Portal Home page - help sections

For registered users, a separate area is available after login, named "*How to*" which has three subsections. One subsection is highlighted in the first column and two other sections are available as links in the second column. Each section in the second column can be maximized in the first column by clicking it:

- A brief overview of ACGT and the introduction to the ACGT Handbook as the main information repository

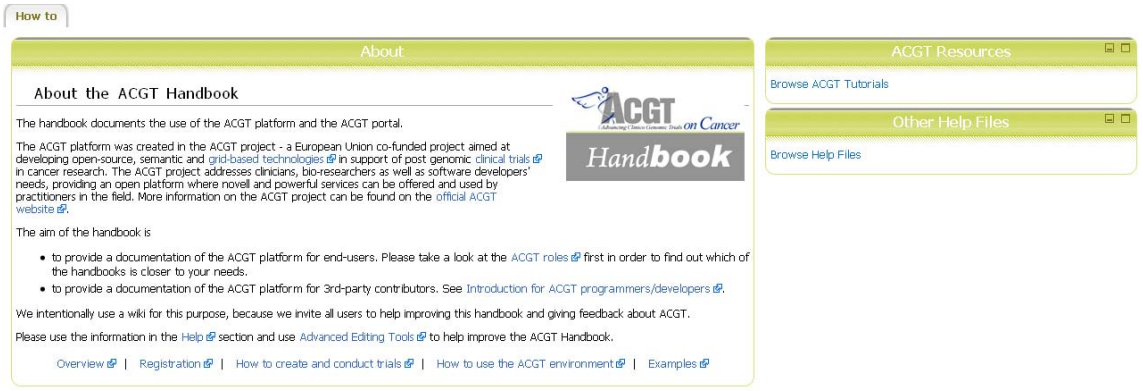

Figure 3 – "How to" – About

- Access to a list of the latest ACGT tutorials. Each tutorial can be selected by clicking the corresponding name in the first sub column. A short preview along with a title and short description are shown in the second sub column. Tutorials can be opened using either the preview image or the link in the "*Link*" section.

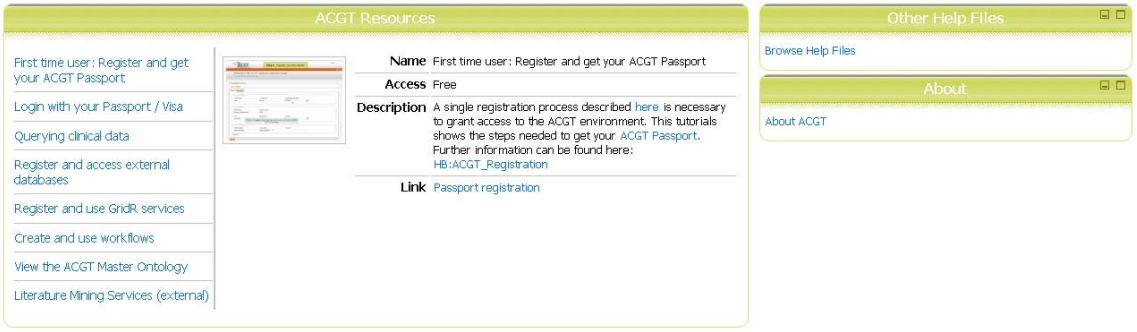

Figure 4 – "How to" – ACGT Resources

PART 2

ACGT Portal prototypes

4 ACGT Portal prototypes

The development of the ACGT Portal was planed to undergo in spiral, based on several prototypes. From this perspective, the ACGT Portal has moved from the first prototype (**P1**), presented in D14.2 to a second prototype (**P2**) (which was scheduled to be presented in deliverable D14.6 before the merge of the deliverables) and a third (and final) prototype (**P3**).

Currently, three versions of the ACGT Portal are maintained – 2 stable versions (running the third prototype) and 1 development version, which contains some ACGT services in an "under-development" state (which can be used in a future prototype).

- **Stable version**, which contains portlets providing access to services that are in a stable development state. This portal is used for demonstrations and user evaluations. This portal can be accessed both using HTTP and HTTPS protocols:

  http://acgt.siveco.ro/ and https://acgt.siveco.ro/

- **Stable version (mirror 1)**, used as a backup in demonstration and user evaluations. The address of this portal is:

  http://epimetheus.ics.forth.gr:8080/

- **Development version**, which contains portlets under development or portlets for services that are still under development. This portal is used for internal consortium-testing only. The address of this portal is:

  http://acgt.siveco.ro:8080/acgt/portal/

The following table shows a list of portlets which are part of the ACGT Portal. Some of these were available in the first prototype and have already been described in the deliverable *D14.2 Visual prototype and report of the ACGT portal*. Others were developed and available in the second and third ACGT Portal prototype.

It should be noted that the second and third ACGT portal prototypes were also focused on improving the usability of the previous portlets and the overall usability of the portal, following evaluation results and user-feedback of the previous prototypes.

Note that portlets label *PX+* have been available since prototype no. X and are still available while portlets marked as *(PX)* have been introduced in prototype no. X and are obsolete.

Portlet (or area) name	***Available since prototype no.***	***Frontend /Backend***	***Short description and comments***
Homepage	P1+	F	Most important changes to this portlet in layout and content are documented in chapter 5 of this deliverable.
Login	P1+	F	Simple login portlet (username/password combination). In P1 using alternative method of direct authentication in the portal (separate database in the portal. In P3 using temporary credentials (Visas) generated by MyProxy Tool.

			Available in the right column of the homepage.
Change password	(P1)	F	Obsolete after P1 (which used Gridsphere Authentication Module), since MyProxy Authentication Module does not allow changing the password used in registration and the temporary password can only be changed in the MyProxyTool.
MyProxyTool	P2+	F	Applet developed by Custodix designed for storing user's grid credentials into MyProxy, wrapped inside a portlet and available in the public area of the portal. Changes from v0.2 (P2) to v1.1 (P3) Available in the right column of the homepage.
Profile manager	P1+	F	Front-end portlet that allows users to change their username, full-name, email address, time zone, language. Can be used as a Group Request portlet, to request access (by emailing administrators) to private groups controlling private areas of the ACGT Portal. Available in the *MySettings / Layout* area in P3.
Role request	P1+	F	Front-end portlet allowing users to request access to new roles (*Join* / *Leave* actions). Available in the *MySettings / Roles* area in P3.
Layout manager	P1+	F	Front-end portlet allowing users to change the overall theme of the portal and to create custom areas in their ACGT workspace. Available in the *MySettings / Layout* area in P3.
My ACGT / Overview portlet	P3+	F	The *My ACGT* area contains an Overview Portlet which has a clickable map of the areas available in the ACGT Portal.
DMS / My Files	P2+	F	The *Data Management* area allows end-users to use the distributed file system of the ACGT Environment. The user interface is called *Data Management System* (DMS) and the corresponding tab in the ACGT Portal toolbar is called *My Files*. It has two separate sub-areas: *File-Directory mode* which allows browsing files and directories

			in the DMS and *Search mode* which allows file searching in the DMS. Available in the *Metadata Registration* area since P2.
Metadata registration	P2+	F	Metadata registration portlets are the user interface to the metadata repository. These portlets provide access to services and data types available in the metadata repository. Available in the *Metadata Registration* area since P2.
GridR Session	P2+	F	The GridR Session area allows end-users to analyze data with R in the ACGT environment. The GridR Session application is used for working with interactive R sessions in ACGT. Available in the *GridR Session* area since P2.
Data Access	P2+	F	The Data Access portlet allows end-users to connect external databases to the ACGT Environment. It also allows custom queries to be executed on the connected databases. Available in the *Data Access* area since P2.
Query and Mapping Tool	P3+	F	The Query Tool is a web-based tool designed to allow end-users to define queries to access the data from clinical trials. Available in the *Query Tool* area since P3.
Workflow editor	P2+	F	The Workflow Editor a web-based tool allowing users to create, edit, execute and publish workflows in the ACGT workflow environment. Available in the *Workflow Editor* area since P2.
Service monitoring	(P2)	F	The service monitoring portlet provides monitoring capabilities for services started by users using the Service Manager portlet. Obsolete and replaced by workflow editor's monitoring area (bottom pane).

			Available in the *MyACGT/ Service Enactings* area since P2.
Ontology	P3+	F	The ontology area allows access to the Ontology Viewer and Ontology Submission Tool. Also provides information regarding ACGT Master Ontology from the ACGT Handbook. Available in the *Ontology* area since P3.
How To	P2+	F	Contains a brief overview of ACGT and the introduction to the ACGT Handbook as the main information repository and a list of the latest ACGT tutorials Available in the *How To* area since P2.
Administration page	P1+	B	General purpose administration portlets that allow members with *ADMIN* Role to configure login options, error handling options (only portal-level errors) and configure chain of authentication modules used. Also shows all current live sessions in a Session Manager and a list of portlet and non-portlet web applications currently deployed and running. Available in the *Administration / Portlets* area in P3.
User manager	P1+	B	Back-end administration portlet - User Account Manager which shows only data from the internal ACGT Portal Database. This internal DB covers custom preference for each Visa (username/ password) that was used to login to the ACGT Portal. Available in the *Administration / Users* area in P3.
Group manager	P1+	B	Back-end administration portlet – Group Manager which allows management of internal portal groups that control access to a list of deployed portlets. Administrators can choose a list of portlets associated with the group (e.g. QueryTool portlet is selected in QueryTool Group) and edit individual user rights. Groups are useful to select global-access rights to portlets based on users' common interests.

			Available in the *Administration* / *Groups* area in P3.
Role manager	P1+	B	Back-end administration portlet – Role Manager which allows management of fine-tuned roles used inside portlets. Roles can be used as a complement of groups to specify access rights to some minor or major functions. E.g. the METADATA_ADMIN role is only used in Metadata Registration portlets to separate users that are allowed to delete metadata entries. Available in the *Administration* / *Roles* area in P3.
Layout	P1+	B	Back-end portlet allowing administrators to change header and footer HTML code, the overall default theme of the portal and guest/group special layouts. Available in the *Administration* / *Layouts* area in P3.
Messaging	P1+	B	Back-end portlet allowing administrators to change mail server configuration information for mails generated by the ACGT Portal (notification emails – e.g. group membership approved, role membership approved). Available in the *Administration* / *Messaging* area in P3.
Credential manager	(P1)	B	Back-end administrative portlet used to retrieve credentials from MyProxy. Its front-end functionality was replaced by MyProxyTool.
GAS	P2+	B	The GAS portlet is an administration portlet application for Gridge Authentication System (GAS). It allows groups, users and object management. Available in the *GAS* area since P2. Requires *jsrgasportlet* group membership which is not default.
GRMS	P2+	B	The Gridge Resource Management System (GRMS) portlet is a client UI for GRMS, which manages the whole process of remote job submission to various batch

			queuing systems, clusters or resources. Available in the *GRMS* area since P2. Requires *grms portlet* group membership which is not default.
VO Management	P3+	B	The **VO Management** is a suite of two web-based tools designed to allow end-users to create their own Virtual Organizations (VOs), to add users and resources to the VO, create roles and assign users to roles inside the VO. Available in the *VO Management* area since P3. Requires *vomanagementportlet* group membership which is not default.

Figure 5 – ACGT Portlets mapped with corresponding ACGT Portal prototype

5 ACGT Portal homepage

5.1 First prototype - homepage

The first prototype's home page reflected the prototype's bottom-up approach focused on the integration of ACGT services. This initial 2-column front-page was minimalistic and deemed too simple and unattractive for new users.

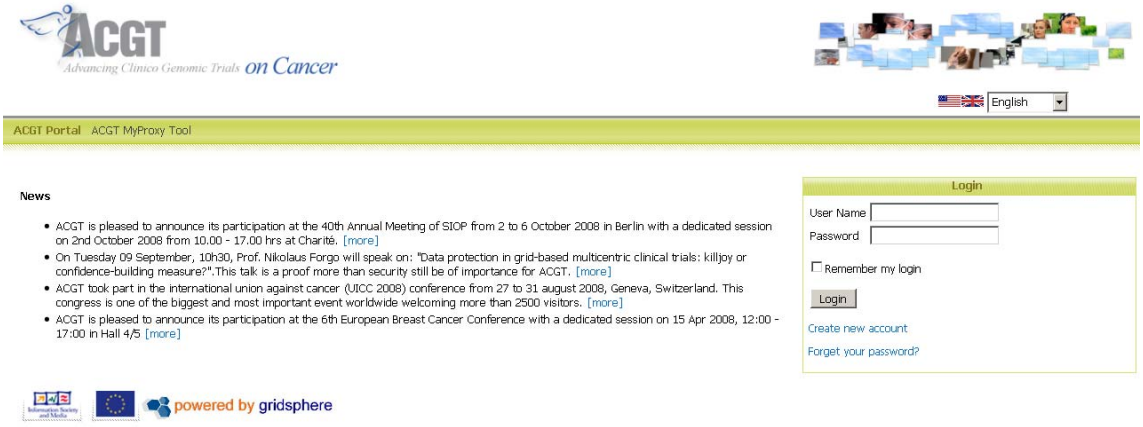

Figure 6 – ACGT Portal homepage – first prototype

5.2 Second prototype - homepage

The top-down approach in the development of the second prototype needed a friendlier ACGT home page in order to facilitate the adoption of first-time users. The new ACGT front-page came with a user-oriented perspective, while keeping the simplicity and usefulness of the design that the user may find in the private sections of the portal.

The home page was designed to provide an attractive, simple entry point that invites potential users, who have never heard of ACGT but stumbled or were referred to this page, to browse and use the applications and available ACGT infrastructure.

The first column (left column) of the homepage was designed to offer a quick glance into the capability of the available applications in the ACGT environment:

- *Top*: shows a current/featured application in a bit more detail and provides a list of available applications.

- *Middle*: Display available video tutorials. This section can accommodate both 'professional' tutorials as videos, but also simple YouTube style and quality videos shot by the users themselves on any subject of interest, after the initial critical mass of users is achieved.

- *Bottom*: a directory of help files arranged by target end-users categories (clinicians, bioinformaticians, etc). Each title is a link to a tutorial or a document.

The second column (right column) was designed to provide hard-to-miss access to registration and login:

- *Top*: a link "Register Now!" to a page with information regarding the registration process which in turn links to the actual registration page (hosted by Custodix, outside the ACGT Portal).
- *Middle*: Login using simplified credentials – replaced by MyProxy Tool v1.0
- *Bottom*: A list of the latest news including new available tools and services, ACGT events, etc.

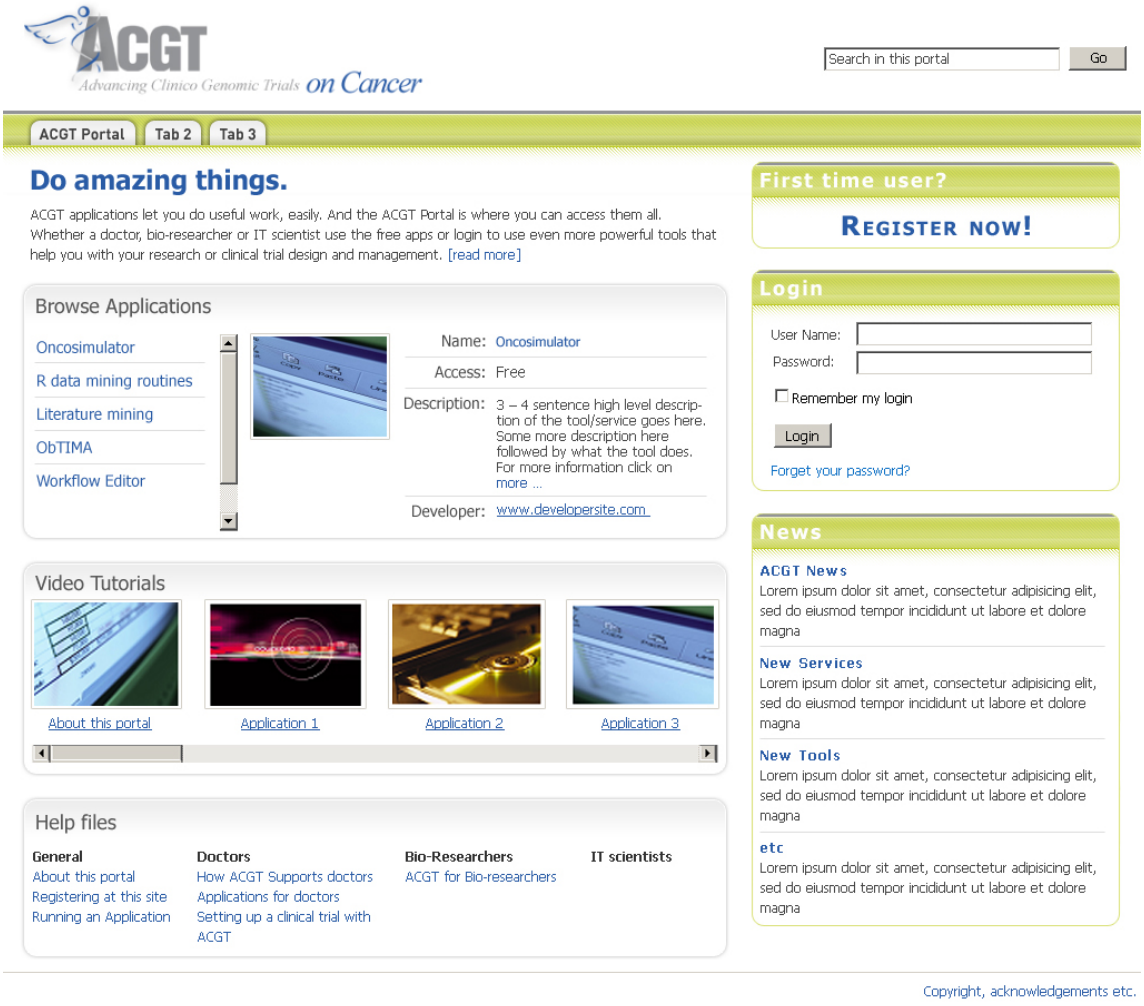

Figure 7 – ACGT Portal homepage – second prototype

5.3 Third prototype - homepage

Following the layout and general design of the second prototype, the third prototype's homepage addressed usability problems regarding registration and login.

Most users in training and evaluation sessions have found the concepts regarding the registration and login to be extremely complicated. The simple login portlet, used as a placeholder in the first prototype had been replaced by *MyProxy tool*, which is a more complicated form of login, capable of authentication using public-key certificates and generating proxy credentials used by heterogeneous services in the ACGT environment.

A new "paradigm"[7] for the registration and login processes was found and implemented in the third prototype – ACGT Passport and ACGT Visa. All related help and tutorial information was updated to use this paradigm.

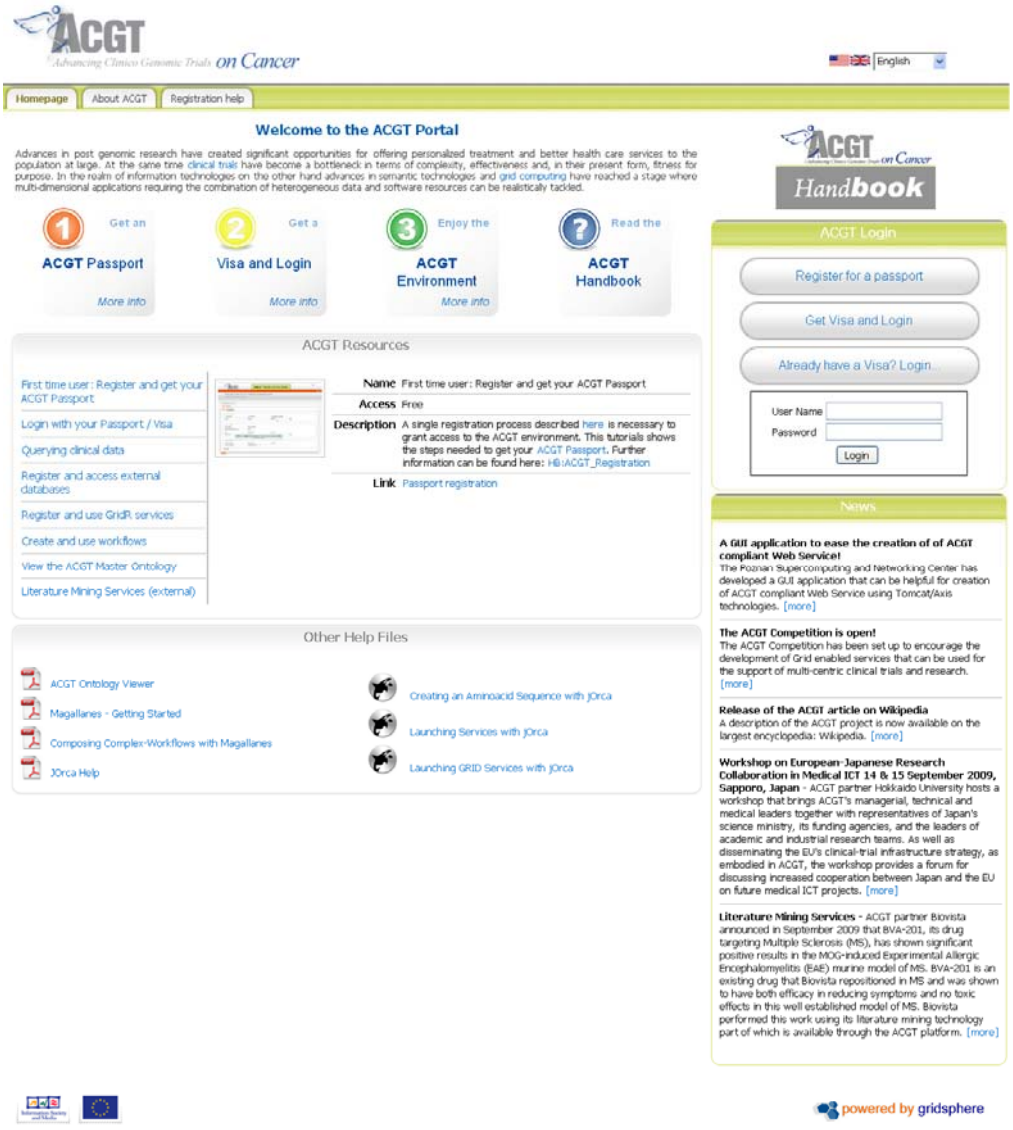

Figure 8 – ACGT Portal homepage – third prototype

[7] see Chapter 6 of this deliverable

The top part of the first column was changed to a horizontal menu with coloured numbered icons - steps for first-time users:

1. Registration (*Get an ACGT Passport*)
2. Login (*Get a Visa*)
3. Explore (*Enjoy the ACGT environment*)

Also, the tutorials are shown using a two-column list as designed in the second prototype.

The top part of the second column was changed to accommodate an animated hard-to-miss link to the ACGT Handbook and a new ACGT Login portlet with three buttons:

- *Register for a passport* - an external link to Custodix's registration form.
- *Get Visa and Login* – which launches the MyProxy applet inside the page, capable of generating short-term credentials (ACGT Visas) in the presence of the ACGT Passport
- *Already have a Visa? Login …* - which contains the simple login portlet from previous versions.

By default, the simple direct login form which uses a Visa (username and password) is shown, as in usual portals.

6 ACGT Passport and ACGT Visas

6.1 Registration and login – history and issues

A simple login portlet using Gridsphere Authentication Module and internal database in the ACGT Portal was used in prototype P1. This portlet was used as a placeholder in prototype P1 and no direct usability was foreseen for other grid-level services until the MyProxy Authentication Module was available. MyProxy is open source software for managing X.509 Public Key Infrastructure (PKI) security credentials (certificates and private keys). A professionally managed MyProxy server provides a more secure storage location for private keys than typical end-user systems and can be used to renew credentials, so that long-running jobs won't fail because of expired credentials.

Access to the MyProxy authentication and authorization mechanisms is provided through the MyProxyTool applet developed by Custodix and wrapped inside a portlet by SIVECO. The ACGT MyProxy installation used passphrases and certificate authentication mechanisms.

These authentication and authorization mechanisms are used in two processes visible to end-users:

- A registration process, after which a public-key certificate is issued by the server, downloaded and stored by the client computer protected by a key-master password known only to the user. This master-password cannot be changed.

- A login process, in which the user uses the public-key certificate to generate MyProxy credentials used for authentication and authorization by ACGT services inside the ACGT environment. These short-term credentials can be chosen - as a pair username / password by the user – and linked to the registered account if only if the user inputs the initial key/master-password (thus confirming ownership of that account). Subsequent authentication requests (after logging in) use these temporary credentials as necessary as long as they are valid. After credentials expire, new credentials can be issued by the user

There are major technical issues concerning security in ACGT (see WP11 deliverables) that have forced the adoption of these complex authentication and authorization mechanisms. However, **most users in training and evaluation sessions have found the concepts and the overall processes of registration and login to be extremely complicated**. More precisely the use of more than one password has created problems for most users.

6.2 The ACGT Passport / Visa paradigm

As technical constraints cannot be changed, to address these issues, a new "paradigm" for the registration and login processes was found and implemented in the third prototype – using two **new** notions: **ACGT Passport** and **ACGT Visa**. The MyProxyTool portlet, the portal homepage and all related help and tutorial information were updated in prototype P3 to use this new paradigm.

Figure 9 – ACGT Passport and ACGT Visa

Main ideas of the new paradigm:

- You need to obtain a passport (once) to be an ACGT citizen – informally: *"be loyal to ACGT rules"*
- Create a Visa for short travel within ACGT boundaries – informally: *"use the ACGT services"*

The overall process can be explained in 4 steps as follows:

Step A. "Register your true identity"

Step B. "Get your ACGT Passport"

Step C. "Register a temporary ACGT Visa"

Step D. "Login in using your Visa"

First, the user's real identity must be confirmed by a Registration Authority, and then an ACGT Passport (a Certificate and a Key Password) will be generated. The Certificate stored on the user's computer and the Key Password chosen during registration acts as a passport in the ACGT environment. The computer (and personal user account) used at step B will act as a permanent residence for the Passport.

The first two steps (A and B) correspond to the registration process and need to be followed once as the ACGT Passport is issued only once. Just like an ordinary passport, the user will be asked for it when crossing the border into the ACGT environment (logging in). However, because of the stronger authorization policies, the user is normally bound to use the ACGT Passport only at the permanent residence - the computer used when it was created.

In order to login from other computers and truly enjoy the diversity of the ACGT environment, the user needs to apply for simple 7 days temporary ACGT Visas - a username and a password. A user can apply for a Visa on the permanent residence computer and then use the Visa on any other computer. When the Visa expires, step C needs to be repeated on the permanent residence computer. The same username can be used all the time when creating Visas as this username is actually an alias which is seen by other ACGT members in the community (instead of your real name). The ACGT Portal automatically matches new usernames to stored preferences under the same passport, so changing the username is transparent to end-users.

6.3 The ACGT Passport explained for end-users

Why is a Passport needed?

Because genuine personal data are flowing in the system, ACGT has a stronger authentication and authorization policy than many online services. The real identity of the users needs to be validated before access to the system is granted. Background information about the legal and ethical approach of ACGT can be found on the ACGT legal and ethical website.

The ACGT Passport is composed of a Certificate and a Key Password received after the confirmation of your real identity by an ACGT Registration Authority (RA), during the registration process.

How do I use the Passport?

The ACGT Passport is stored on the computer you used in the registration process and can be used in the login process **only from that computer**.

To login from a different computer, you will need to:

- Transfer the ACGT passport from that computer to a USB stick and insert the stick in other computers in the login process. See this section Transferring the ACGT Passport for more information.

OR

- Apply for 7-day-limited ACGT Visas which can be used from any computer. See this section Login using ACGT Visas for more information.

The ACGT Passport & Password explained

Your identity (or account) is composed of a digital **Passport** stored on the computer and a **Password to unlock the Passport** which you choose and remember in the registration process.

The two separate components are 2 different secrets needed for your identity to be acknowledged in the ACGT environment:

- *something that you need to have* - the Certificate/Passport
- *something that you need to know* - the Password required to *unlock* the Certificate.

Caution:The Key Passport cannot be modified after the registration process.

In turn, the ACGT Passport has two separate electronic components, stored locally on your computer:

- a Public Key Certificate
- a Private Key.

The Passport uses **Public Key Cryptography** which is based on a pair of two asymmetrical keys: a public key and a private key. The private key is kept secret and the public key is widely distributed.

The secret of the Passport - its **private key** - is protected by a regular human-readable password. You choose this password during the registration process and you need to remember it to use the ACGT Passport.

Caution: In the **Login** section, when you get your Visa and login, you will need this password. It's called "**Password to unlock my Passport**" on the login interface.

Note: It is formally called the **Key Password** because it protects the key used to protect the certificate. See the example key in the following section.

Your ACGT Public Key Certificate

Your Certificate contains your Public Key, which is generated in **step B**. A Certificate Authority (the Registration Authority in our case) certifies that some identity information (that you provided in **Step A**) and your Public Key belong together. This digital signature confirms your identity and is required whenever you want to login to the ACGT environment.

The certificate is stored in a file called *usercert.pem* in the *.globus* directory.

Example:

-----BEGIN CERTIFICATE-----
MIIDvjCCAqagAwIBAgIIKGyUMs9kTdMwDQYJKoZIhvcNAQEFBQAwSTELMAkGA1UE
BgwCQkUxETAPBgNVBAoMCEN1c3RvZGl4MRAwDgYDVQQLDAdQcmltYXJ5MRUwEwYD
VQQDDAxBQ0dUIENBIDIwMDYwHhcNMDkwNjE3MTMwNjI1WhcNMTQwNTIyMTMxNjI1

WjBKMQswCQYDVQQGDAJFVTENMAsGA1UECgwEQUNHVDEPMA0GA1UECwwGU0lWRUNP
MRswGQYDVQQDDBJBbGV4YW5kcnUgU29mcm9uaWEwgZ8wDQYJKoZIhvcNAQEBBQAD
gY0AMIGJAoGBAIjzzPxYthPoY871JyNqnmsrzakc864wJTxsTygpuW04Od1OFCS0
G3vhEmq2StvEEbCCxfjxrwYQo0d7d1XPYfh9X5jZScuNGUbTKzZb2N9lTvvO8nSF
BkREkJhqrtXfTzugKd8TSM0piczYpk2wQ9ITuYXJqWLv+XC9Q+3V+lVDAgMBAAGj
ggErMIIBJzAMBgNVHRMBAf8EAjAAMA4GA1UdDwEB/wQEAwICpDAdBgNVHQ4EFgQU
H42aZKj+kRk9tswGBg1Fvp1g8MMwHwYDVR0jBBgwFoAUBOR4Y+LKgslzYVoBCJ4F
0DgBE/EwGAYDVR0gBBEwDzANBgsrBgEEAc1yAQEGBTBqBgNVHR8EYzBhMF+gXaBb
hllodHRwOi8vY2EuY3VzdG9kaXguY29tL2NybC9jcmxTdG9yZT9pc3N1ZXI9Y249
QUNHVCUyMENBJTIwMjAwNixvdT1QcmltYXJ5LG89Q3VzdG9kaXgsYz1CRTBBBggr
BgEFBQcBAQQ1MDMwMQYIKwYBBQUHMAGGJWh0dHA6Ly9jYS5jdXN0b2RpeC5jb20v
b2NzcC9yZXNwb25kZXIwDQYJKoZIhvcNAQEFBQADggEBADwwv+K1GBePPXqQ7NjI
IOlwMrxkk9yjuavaORIDU4pgo+wqbKgJwbCqAKCeZVLAjKKxX389JwiLloNm7ziX
PidZVqjPOwtmFW7YIigwJrHOfV58UQdlWeg/j0a3CBP+haBEjAASfIBXcaN80Je/
SopJ+xxoWOyenmWfLoG8373VmPky3XgRSndoRqBvIyb8sesJ7Xq3xkOApMpyG3B7
KQoOPivC9Tq2UVLgISx49M3pDnvPGbPBdBp2YWGp9iglviXMyru2CtE3iD0ng7g+
q9weDlUSnafUnR8KXdOCsHJIVlFSBkEgSkmfO9mkS1yoj0jo1uL+Jfk51PSgox09
/Fs=
-----END CERTIFICATE-----

Your ACGT Private Key

Your ACGT Private Key is the pair of the public key inside your ACGT Certificate. The login process requires confirming the identity stored in the ACGT Certificate, which in turn requires proving you possess the private key that pairs with the private key contained by the certificate.

In order to protect your Private Key, it is encrypted in **Step B** with the "Key Password" you choose. The private key is stored only encrypted and does not leave your computer during the registration process.

Note that this is done with a Symmetric Key Algorithm so there is only **one** key this time - the Key Password.

The private key is stored in a file called *userkey.pem* in the *.globus* directory. Example:

-----BEGIN RSA PRIVATE KEY-----
Proc-Type: 4,ENCRYPTED
DEK-Info: DES-EDE3-CBC,9123ca4cfb2d6909

64Jz1lwQq7gQwAfJ5SbKg2Kfdskxp6s7p+bi5iIJVFeS7xdQJoly/FyAguvnJDJe
ur/+osVHjaAZ/dMe2MQvKy6vcryphP86bdJ7C312ws0FCj10jQNuj74QQztZr3aD
XnBsgCIkaHPx/8DbDatncR/OJjZqkFW6D4gxyQ64uop0/ffEvx77M8AN0xqYbkPI
X/vzPeu80FNukuiyEeRQZm9Jpc+lNv7WpmcqLcsDYB05Rx8t2EYzB1DdHfQK8qux
qSlZhfynGumQxD2OxE87sDFuTjD+4Twg/KQC4yhxcS5GjG4sXR+/B9O+fnvnP+GM
2FaT7AO3KEAy4FuquHUNCkDo39vuvezpILbvC8rFjT+fAwyWeLB/gRRP9EtdFZZE
JCiggsOvk2P8FeldFKeR2KVsFdmeG8t/Mr/rxTuXADk2+9T+fMkUzbySo5ZXabO+
YRxHiALO+T3HYHowxCnnuFj2H1f/XcDUh7Zc5VQ77PlZHWMXsF50T1F8uXslkgj+
eQsoIMi2+VBJDifC9PhVcn4miG6UY2SQXQtj/RYSvSiRdbHyP7j0lKCfZu9j7S24
LBV8ce3Z37qtWJCz55DyE3rtRgWEqDX/iE7NEq5vXutzzONpDN8HJzAV52t/rgHl
kwNp/xTrdx91yOXAVPfB+3kg6/I9MZJUb2M79wmSD/7uvGFTGJyki8Un6JSyuHvH
mEggNsITDci0QqBUKAFtNPyN1ryoLFd5rtixJ5g3F/dGxl0BsqZ/XwpifklrB1Zu
mXTDZf7OQTo4j8KZBj7BtdhZJmZ0g2f2I3dwduLrsk8=
-----END RSA PRIVATE KEY-----

6.4 Detailed registration and login explanations for end-users

Step A. "Register your true identity"

First, your identity must be verified and registered with an ACGT Registration Authority. Go to the registration site (using the **Register for a passport** section on the ACGT Portal homepage) and fill in the required fields. You must select the organization to which you belong as the Registration Authority capable of verifying your data. Please enter a valid email address necessary for the following steps. The ACGT Registration Authority (RA) will review your application and if necessary contact you for further information or instructions. Wait for these instructions from the RA.

You can use this time to make some configurations to your browser: check and install Java Runtime Environment and add the ACGT Certificate Authority to your trusted Certificate Authorities. Following these steps will ensure that your browser will communicate with the applications in the ACGT Environment which are certified by the ACGT CA - this step is optional, you can choose to individually trust each of these applications, when you try to access each one in the ACGT environment.

Step B. "Get your ACGT Passport"

If your application is approved you'll receive an e-mail with an activation link. **Please complete the following steps on the computer and personal user account you will use** to access ACGT environment. Your ACGT Passport will be issued at this step and it will be stored on the computer that you use for the following steps.

Note: You will need to have Java installed in your Browser for the following steps. Please check the Java Runtime Environment version you have installed at http://www.javatester.org/version.html. If you do not have Java installed or your version

is lower than 1.5, you will have to download the latest version of Java Runtime Environment at http://java.sun.com/javase/downloads/

Click on the link you received to start the generation of your ACGT Certificate. The automated process will create your certificate based on your personal data you provided at step A. The certificate will be installed on your computer and can only be used by the personal user account you use during the creation of the certificate (it cannot be used by all the users accessing that computer).

Caution: As a security precaution, you should generate the certificate on your private computer (where you are the only person who uses the computer) using your own personal account (instead of a "Guest" account).

To protect your certificate, you have to choose and remember a Password. You will be asked for this password to "unlock" your ACGT Passport - in order to login the ACGT environment by generating ACGT Visas, which are needed to login from other computers. Once a Visa is generated, you will not need the Passport for 7 days.

See *"ACGT Password explained"* for an explanation of the components and the reasons for the ACGT Passport.

Step C. "Register a temporary ACGT Visa"

Use the **Get Visa and Login** section in the right panel of the ACGT Portal to generate a temporary 7 day Visa for the ACGT environment. After clicking Login a dialog will pop up with a login form. This step requires that the ACGT Certificate from your ACGT Passport is present on the computer (or on a USB stick connected to your computer). You will see the identity from your ACGT Passport already selected on the first row of the form. Type your password - the one you chose when your passport was issued - to unlock and prove that you are the rightful owner of the ACGT Passport. Next, you create a new Visa by choosing a username (alias) and enter a new temporary password which can be used for 7 days. Alternatively, you can choose to remember this Visa so that it recreates itself each time you login from your computer. After creating the new Visa, the system will automatically log you on. You can use the Visa for 7 days from any computer without the need for the ACGT Passport (See Step D for more details).

Note: Each time you get a Visa you can change the username (alias) and the temporary (Visa) password. Also, please note that for increased security, this temporary Visa password is required to be more complex: it must have at least 7 characters and contain both letters and numbers! (MyProxy v1.1+)

Caution: Step C is **the recommended way** of logging in the ACGT environment. It is the most secure as it always requires knowledge of the password which protects your Passport. Since step C effectively logs you in, you will not need to go through step D as long as you use a single computer or you transfer your ACGT Passport to a USB Stick and attach it to another computer you want to use.

Note: The ACGT Passport is stored in a directory called /.globus in your personal home directory on the computer you used in Step B. This directory can be copied to a USB stick and used to login in from other computers.

Step D. "Login in using your Visa"

It is recommended that you transfer your ACGT Passport to a USB Stick and use it to log in from other computers using Step C. Step D should only be used when you need to access the ACGT environment on other computers and you do not have access to your ACGT Passport.

Use the "**Already have a Visa? Login..."** section in the right panel of the ACGT Portal to log in the ACGT environment. Please use the temporary Visa - username and password you have generated in Step C - to login to the ACGT environment.

Remember that a Visa only lasts 7 days and you will need access to your ACGT Passport in order to generate another Visa.

7 ACGT Front-end Portlets

7.1 MyProxyTool Portlet

MyProxy is open source software for managing X.509 Public Key Infrastructure (PKI) security credentials (certificates and private keys). MyProxy combines an online credential repository with an online certificate authority to allow users to securely obtain credentials when and where needed. Users run myproxy-logon to authenticate and obtain credentials, including trusted CA certificates and Certificate Revocation Lists (CRLs).

Storing credentials in a MyProxy repository allows users to easily obtain RFC 3820 proxy credentials, without worrying about managing private key and certificate files. They can use MyProxy to delegate credentials to services acting on their behalf (like a grid portal) by storing credentials in the MyProxy repository and sending the MyProxy passphrase to the service. They can also use MyProxy to renew their credentials, so, for example, long-running jobs don't fail because of expired credentials. A professionally managed MyProxy server can provide a more secure storage location for private keys than typical end-user systems. MyProxy can be configured to encrypt all private keys in the repository with user-chosen passphrases, with server-enforced policies for passphrase quality. By using a proxy credential delegation protocol, MyProxy allows users to obtain proxy credentials when needed without ever transferring private keys over the network.

MyProxy provides a set of flexible authentication and authorization mechanisms for controlling access to credentials. Server-wide policies allow the MyProxy administrator to control how credentials may be used. Per-credential policies provide additional controls for credential owners. MyProxy supports multiple authentication mechanisms, including passphrase, certificate, Kerberos, Pubcookie, VOMS, PAM, LDAP, SASL and OTP.

In the ACGT Portal, MyProxyTool is an applet developed by Custodix designed for storing user's grid credentials into MyProxy. The applet was wrapped inside a portlet by SIVECO and deployed on the ACGT Portal Homepage.

The first deployed version of the MyProxyTool (v0.2 in P2) had a simplified interface which allowed users to delegate credentials and remove delegated credentials.

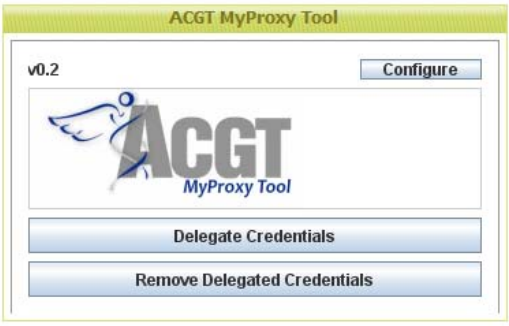

Figure 10 – MyProxy Tool Portlet v0.2 (P2)

The final version of the MyProxyTool (v1.1 in P3) was embedded in a different portlet, according to the new ACGT Passport/Visa paradigm (see chapter 6). In this new paradigm the MyProxyTool is used to issue Visas to Passport holders in the login process.

Figure 11 – MyProxy Tool Portlet v1.1 (P3)

The applet's login user interface has been modified in v1.1 by Custodix and SIVECO to reflect the changes. There are two marked areas "*I have my ACGT Passport ready*" and "*Use this ACGT Visa*". Two passwords are still needed, but are clearly marked as "*Password to unlock my Passport*" and "*ACGT Portal password for new 7-day Visa*" respectively. As the applet remembers the Visa username and password and has an option to automatically reuse the locally stored Visa password, the new applet can effectively be used with a single password (the Passport's password) every time.

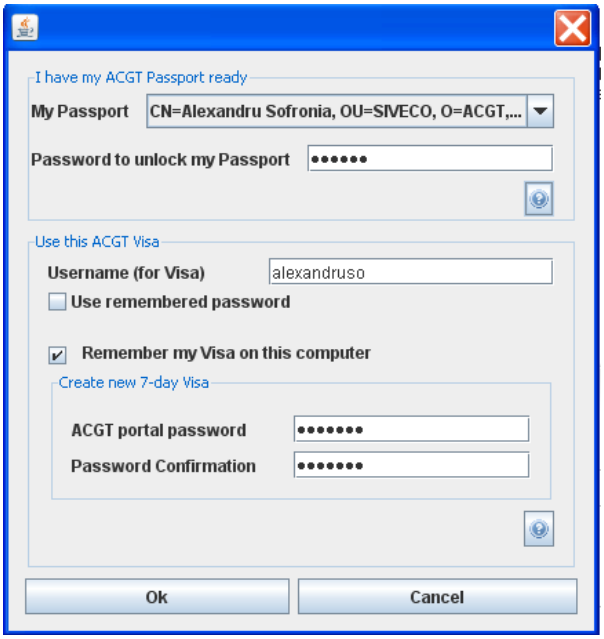

Figure 12 – MyProxy Tool login interface v1.1 (P3)

Also, embedded help from the ACGT Handbook, for both the ACGT Passport and the ACGT Visa, is available as question mark buttons in the applet. More information is available in the ACGT Handbook http://handbook.eu-acgt.org/HB:Logging_in_and_out.

A step-by-step tutorial is available in the ACGT Portal, named ACGT Login

7.2 MySettings

The *MySettings* area of the ACGT Portal allows users to change their profile, roles and group membership and customize the ACGT Portal Look & Feel.

Figure 13 – MySettings area (P1)

7.2.1 Profile Manager

The *Profile* section allows users to change their profile settings and group membership. Users can change their Full Name, organization, email address, locale (English is default) and time zone information. All changed information is kept in the portal's internal database as stored user preferences and has no influence on grid-level. For example, the key password (that protects one's Passport) cannot be changed from the portal.

Users can choose their group membership to public and private groups. Any *public group* is joined automatically by ticking the appropriate box. Joining a *private group* needs the administrator's approval. The description of a private group ends with the text "*This group is private. Please email the administrator"* with a *"mailto:"* link on the last word which automatically completes the subject of the email with the actual group name: "*mailto:acgtsupport@siveco.ro?SUBJECT=[ACGT] Group Join Request: ActualGroupName"*.

E.g. the GridR Session portlet is only visible to the users belonging to the *gridrsession* group. This is a public group which can be joined and left by any user. These leave/join actions effectively toggle the *GridRSession* tab in the ACGT Portal upper menu for the user.

Note that the mails requesting group membership do not use the internal messaging functions of the portal, as role requests do.

The following groups are available and grant access to the corresponding portlets and ACGT sections (tabs). All public groups are automatically selected for new users:

- *dms* - Data Management System portlets (My Files area)
- *wegroup* - Workflow Editor portlets
- *metadataregistration2* - Metadata Registration portlets
- *querytool* - Query Tool portlets
- *gridrsession* - GridR Session portlets
- *vomanagementportlet* - VO Management portlets - **Private Group**

- *dataaccess* - External Data Access portlets (Data Access area)

7.2.2 Role Request

Specific ACGT roles give various rights in the ACGT Portal, e.g. management of access to data pools and to various elements of the interface. The Role request portlet allows users to request to obtain new roles by clicking the **Join** button next to the desired role. An email with this request will be sent to the ACGT Portal administrators through the internal messaging system of the ACGT Portal. Users can give up a role by clicking the **Leave** button next to the desired role and the modification will be automatically enforced.

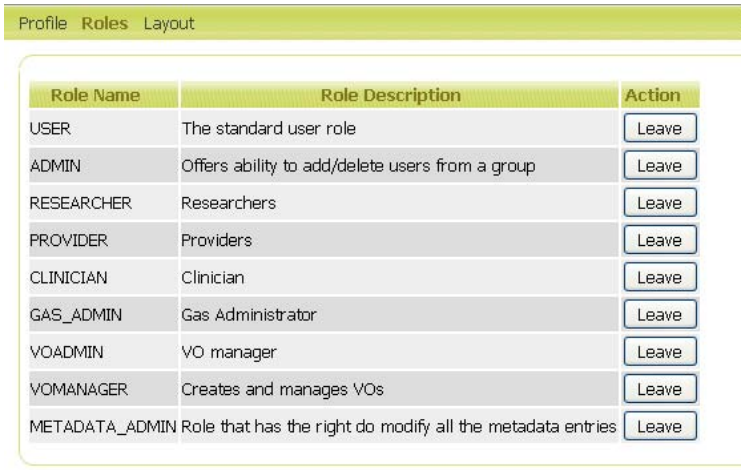

Figure 14 – Role Request (P1)

Roles and groups in the ACGT Portal are both used for access rights in different ways. For example, the *metadataregistration2* group defines which users are allowed to access the Metadata Registration portlet, while the METADATA_ADMIN role defines which users are allowed to modify the metadata entries. In this example, the group membership is used to grant access to one or more portlets, while the role can potentially be used in all portlets to grant specific rights.

7.2.3 Layout Manager

Figure 15 – Layout Manager (P1)

The *Layout Manager* in the *MySettings* area can be used to change the theme (skin) for the ACGT portal or further customize the ACGT toolbar (menu) and workspace by adding new tabs. All changes are saved in the internal portal database and only affect the current user.

New tabs in the menu (or areas) can be created so that one can combine existing tools and resources (portlets) to create customized pages.

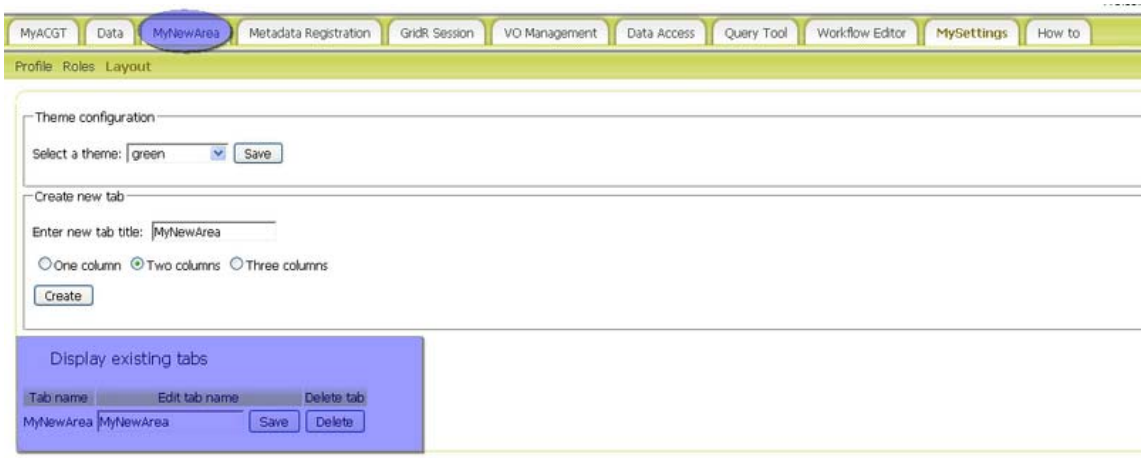

Figure 16 – Custom tabs in the ACGT Portal (P1)

For example, one could create a new 1-column area which shows the *Data access* portlet and the *Workflow Editor*, instead of using the two default different pages for workflows. A user might find using the two portlets on the same page to be much easier when designing workflows that require a large number of external data.

7.3 My ACGT

The *My ACGT* area contains an *Overview Portlet* which has a clickable map of the areas available in the ACGT Portal. The Overview Portlet was derived from earlier PPT presentations to provide a clickable task-oriented map of the areas in the ACGT Portal, complementary to the 1-dimensional menu bar on top of the ACGT Portal. This area is the default landing page after logging in, so it provides an overview guide for both first time and regular users.

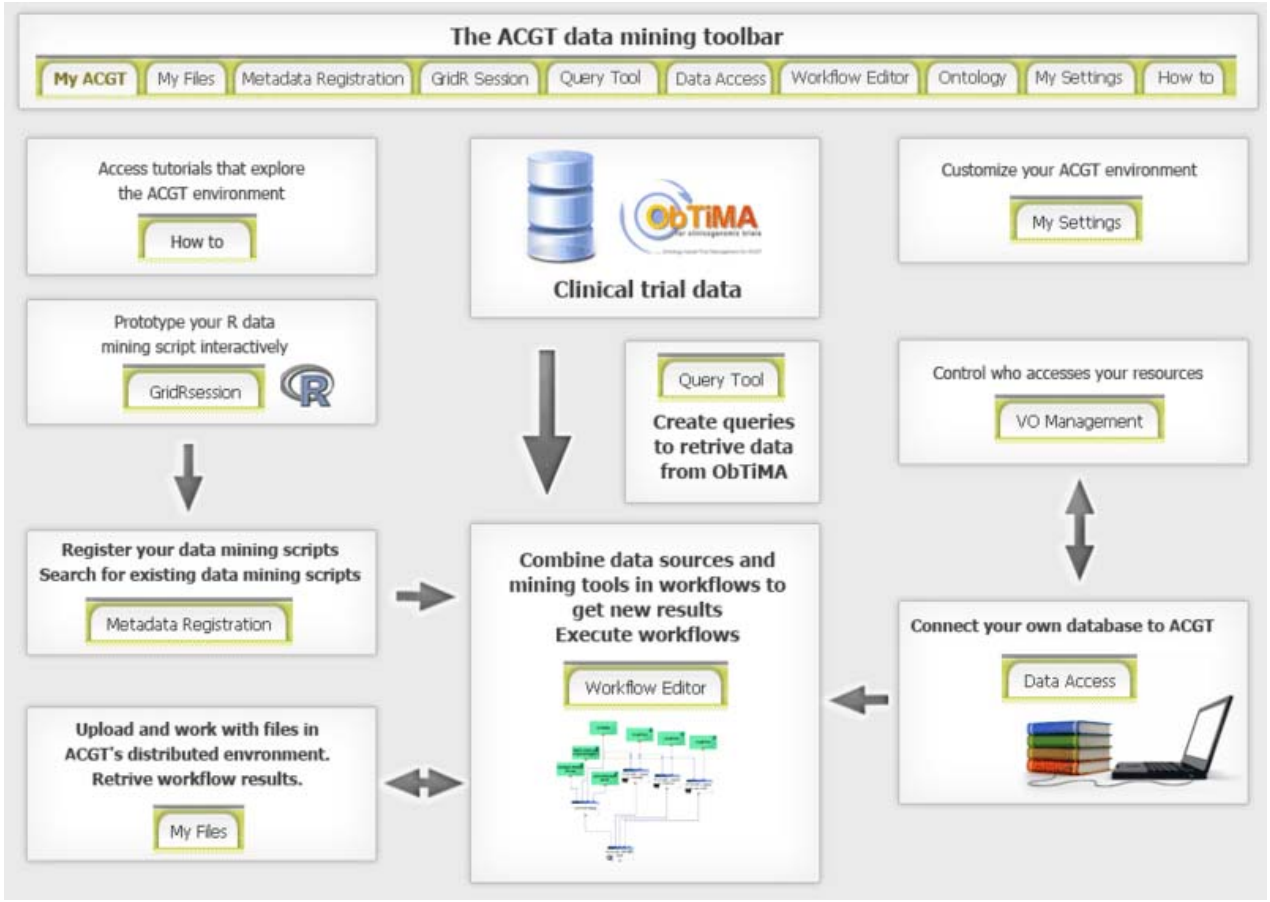

Figure 17 – ACGT Overview Portlet (P3)

7.4 DMS Portlet

Gridge Data Management System is one of the main components of Gridge Data Management Suite (GDMSuite) - a middleware platform providing a uniform interface for connecting heterogeneous data sources over a network. GDMSuite stands for the backbone of the Gridge environment, on which computational services would perform its operations. Gridge Data Management Suite constitutes a bundle of packages, designed for the creation of a complete and robust data management environment. It is intended to fulfill even the enterprise requirements of grid environments in terms of reliability, security and performance.

Data Management System (DMS) is a middleware application developed by PSNC, based on SOA model that determines a loose coupling between reusable components, which are platform-, language-, and operating system-independent. Similar to computing and network resources, DMS provides services to manage and retrieve data files in order to support grid jobs. The computational resources managed by DMS can be described by metadata scheme, which allows create an abstract, semantic and explorable layer of resources.

Originally, DMS was provided with a stand-alone Web interface called DMS Portal. The DMS Portlet was developed by SIVECO by emulating all the features available in the DMS Portal and introducing several value-added features. Also the web service layer used by the DMS Portal to access the backend applications was ported into the DMS Portlet.

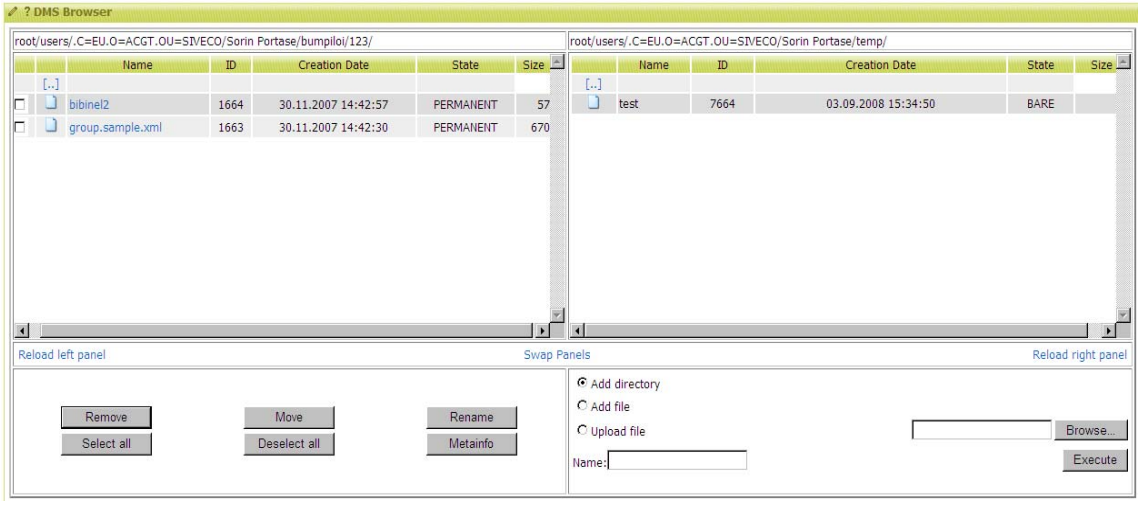

Figure 18 – DMS Portlet (P2)

The DMS Portlet has two different view modes available:

- The **File-Directory mode** which allows users to browse files and directories in the DMS.
- The **Search mode** which allows users to search for files in the DMS.

The DMS Portlet has been continuously upgraded, both in an effort to solve usability problems and add user-requested features, and in an effort to enhance its performance using improved versions of the underlying low-level functions (from PSNC). The DMS version in prototype 3 is slightly different than the initial version in prototype 2.

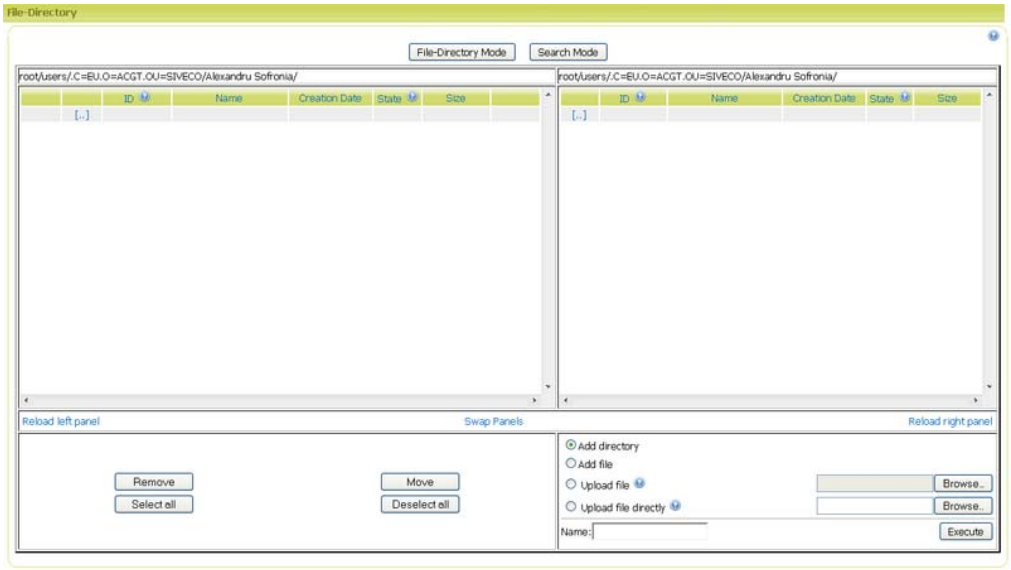

Figure 19 – DMS Portlet v2. 0 (P3)

7.4.1 File-Directory Mode

The **File-Directory mode** shows two panels *(Left and Right Panel)* on top of another panel *(Action Panel)* with *actions* that can be performed.

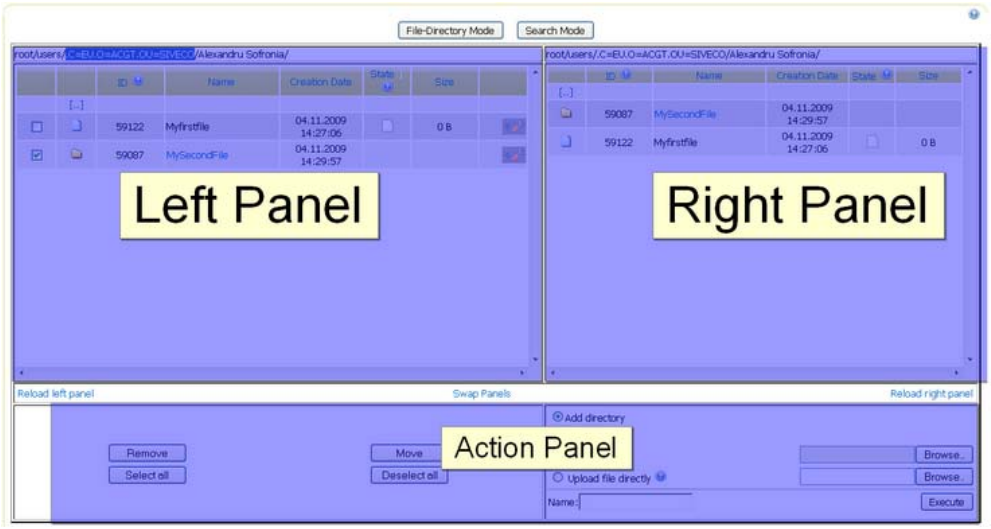

Figure 20 – DMS File-Directory Panels (P3)

- The **Left and Right Panels** allow navigation through files and directories in the DMS. The two panels are **independent** and can be **refreshed** (or reloaded) independently to see the new status of the files. S**wapping** between the left and right panel is possible.
- The **Action Panel** groups the various actions that can be performed in the DMS File System.

Each **panel** has a header with 8 columns:

- Select files/directories using checkboxes (in order to perform the same action on multiple files/directories at once)
- **Type** - file/directory
- **ID** - Unique Identifier of the File - generated by the system. In the ACGT environment files and directories are internally identified by their **Unique File Identifier** (a numerical constant) and **not** by their File Name. So it is possible (though **not recommended**) to have several files with the **same name** in the same directory.
- **Name** - Name of the file as specified in the *adding* or *uploading* process.
- **Creation Date** - date of the creation of the file (regardless of how it was added or uploaded).
- **State** - one of the 3 states a file can have in a distributed file system.
- **Size** - size of the file

Each column is sortable by creation date, site and status simply by clicking the corresponding column headers.

Above each panel the **current working directory** is displayed (the path of the directory in the DMS file system). Default is the personal workspace: *root/users/ ORGANIZATION/ USER-NAME*

7.4.2 File states in the DMS

When a user or one of the processes that run in the ACGT environment create, add or upload files to the DMS, there are 3 possible states in which a file can be. A file can be in various states during the execution of a process - *bare* / *in progress* / *permanent*. A file can be marked "*in progress*":

- When a file is created or uploaded in a grid environment, most often the file is uploaded through the ACGT portal (one machine) to another machine in the grid environment. This implies that it is "in transit" for a while, depending on the connection speed between all these machines (your computer, ACGT Portal, final destination in the grid)
- When a process in the ACGT environment, running on *machine A*, finishes its execution and tries to save its data, it uses a file in the DMS which may sometime be stored on a different *machine B*. During this saving process, the file is in transit.

Only the **permanent** state corresponds to a fixed final state. The other two states (*bare* and *in progress*) can change at any moment, without prior notice!

- In the **BARE** state the file has been created (either by a user or after a process starts execution) but it has not received **yet** any contents. It is empty (**0 B**).
- In the **IN_PROGRESS** state:
  - A file is uploaded from a local computer through the ACGT Portal to another machine in the DMS - so the content is still **in transit**, or
  - The process that created or is using the file has finished its execution and is saving its data to the file in the DMS.

- In the **PERMANENT** state, the file content is stable as it has been successfully transferred to the DMS (either through the actions of a user or a process in the environment).

Files remaining in **BARE** or **IN_PROGRESS** state after a process has completed indicate a problem. Only files in PERMANENT state contain valid information that can be safely used!

7.4.3 Action Panel

Various actions can be done in the action panel:

- **Add a sub-directory** to the current working directory:
- **Upload** a file to the current working directory (2 options):
    - *Upload a file through the portal*. The file will be uploaded into the ACGT portal machine and then it will be transferred to the storage location. This is very useful when trying to upload small files as it does not require a direct connection to the storage machine in the grid.
    - *Upload a file directly*. The file will be sent directly to the storage location in the grid. This allows direct upload of larger files but it requires direct access to the storage machine in the grid.
- **Delete** files and directories: Check-boxes corresponding to the files and directories to be deleted must be checked in the left or right panel, and then the *Remove* button in the Action panel deletes them from the DMS.
- **Move** files and directories: First navigation using the left and right panels must ensure that the source directory is located in the left panel and the target directory in the right panel. In the left panel the check-boxes corresponding to the files and directories must be checked, and then the *Move* button will start the actual process.

7.4.4 Action Menu

Each file or directory displayed in the left panel has its own menu in the last column which allows various actions on the selected file or directory:

- **Rename** a file or a directory. A dialog will appear to fill in the new name of the file/directory.
- **Manage the meta-data** information associated to a file or a directory - "Meta info" option. The meta-data information will be displayed into the right panel and can be edited separately.
- **Manage file permissions** - "File permissions" option. The permission management interface will be displayed in the right panel. There are 4 usual permissions (Read, Write, eXecute, Permissions write) and any user or group can be selected from the internal portal database.

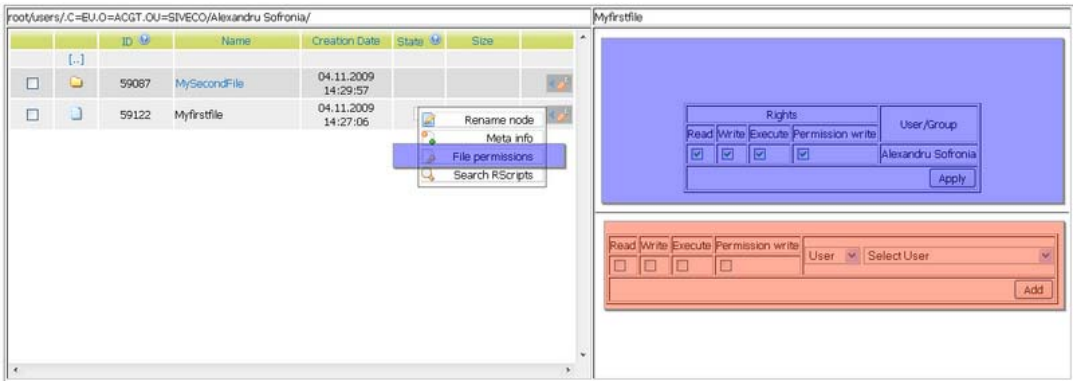

Figure 21 – DMS – File Permissions (P3)

- **Search for** compatible R-based **services** - "Search RScripts" option. The application will display a list of R-based services compatible with the type of the given file.
- **Download** - "Download" option. A dialog will appear to choose where to save the file on the local computer.

7.4.5 Search Mode

The **Search Mode** allows searching for a file using various criteria:

- Described by file attributes
- Described by file metadata schema
- Described by file metadata attributes

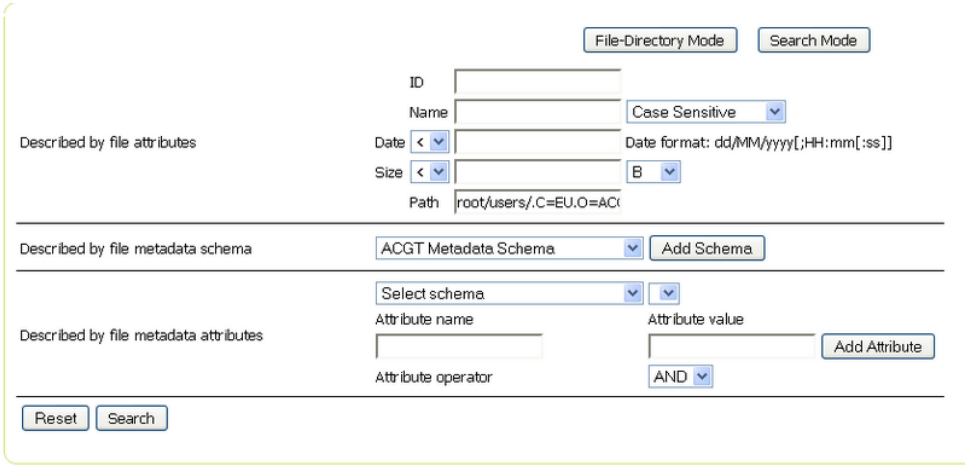

Figure 22 – DMS Search Mode (P3)

More information on the DMS Portlet can be found in the Handbook: *HB:Data_Management* and *HB:Storing_data*.

7.5 Metadata Registration Portlets

The user interface to the Metadata Repository is a JSR 168 compliant portlet application that allows a user to browse and edit the Services and Data types stored into the Metadata Repository. The Metadata Registration Portlets v2.0 were developed using available Gridsphere tools and relay on plain JSP and minimal JavaScript for rendering. These portlets replace the Metadata Registration Portlets designed for the first prototype of the ACGT Portal and presented in D14.2.

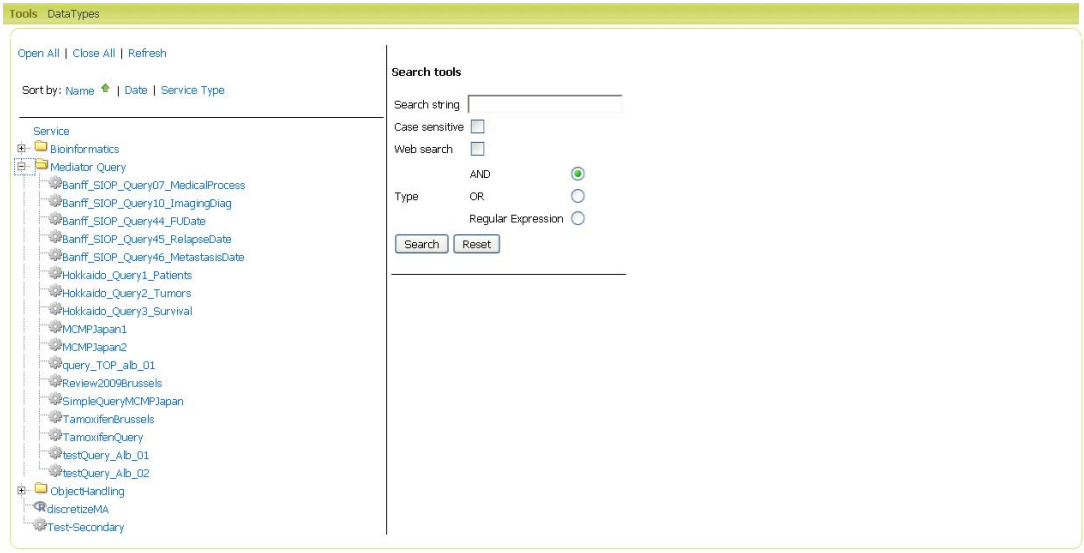

Figure 23 – Metadata Registration Portlet v2.0 (P2)

Browsing the existing services as well as registering new services is done through the **Metadata Registration** tab of the portal, selecting the **Tools** sub-tab.

- The tree on the left hand side can be used to browse the services classified in various functional categories.
- Services can be searched using keywords in the **search** tool in the upper-right side of the portlet.

Registering a new service (or creating a new functional (sub)category) in a specific functional category can be done by clicking on the category in the tree on the left, then by selecting the specific action to be done in the lower right part of the interface.

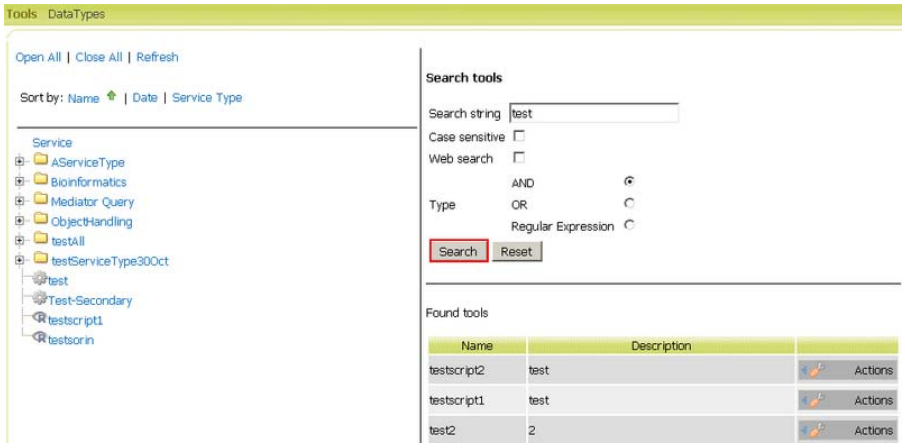

Figure 24 – Metadata Registration Portlet – Searching tools and services

The interface provides an easy to use search engine for tools/R-based services. The search can be case sensitive, or case insensitive, depending on the user's selection. To use this engine, the user must fill in the search string and provide the search rule:

1. a combination of all the words contained in the search string
2. at least one of the words contained in the search string
3. a regular expression

Browsing and registration of data types is based on the same principles as for services, by selecting the "*DataTypes*" sub tab of the "*Metadata Registration*" tab of the portal.

Clicking on the "*Add new data type*" link opens a new interface which can be used to register a data type. A number of data type-description fields can then be filled. If the user wants to create a non primitive data type, he can use the existing data types to compose its new data type by clicking "*Add data type component*" link and selecting a relation type, and a data type the relation refers to.

Note that when creating a sub data-type (by selecting a data type and then clicking the "*Add new data type*" button) an "*Extended*" relation is automatically created between the parent data type and the newly created data type. After the data type has been created, one can select it from the data type tree and modify its attributes, or delete it.

The data type interface provides an easy to use search engine for data types. To use this engine, the user must fill in the search string and provide the search rule:

1. a combination of all the words contained in the search string
2. at least one of the words contained in the search string
3. a regular expression

More information on the Metadata Registration Portlet can be found in the Handbook: *Registering services and Registering data types.* A sample tutorial is available on the ACGT Portal, named Register and use GridR services.

7.6 GridR Session Portlet

The GridR Session is an applet developed by FhG as a GridR web-based client. The applet was wrapped inside a portlet available in the *GridR Session* area since P2.

In ACGT the R environment, namely GridR, is used as a tool for the remote execution of R code in the grid. More specifically, the task of the execution of the R code is submitted as a grid job to a remote grid machine. The current implementation of the server side GridR components that are related to the grid environment is based on several external software components, namely the GT4 grid middleware, an installation of the R environment on the grid machines which will execute the functions remotely and a GRMS-Server installation from the Gridge toolkit on a central machine in the grid environment that is responsible, for instance, for resource management.

On the client side, GridR consists of a set of R functions and involves the Cogkit, which is responsible for proxy generation and data transfer, and a GRMS-Client. The client side part is structured around the components "RemoteExecution" (JobSubmission and JobDescription Generator) and "Locking". The RemoteExecution component is responsible for the execution of R code as a job in the grid environment by transforming the R code to execute into a set of files, creating a job description file in the respective job description language, and submitting the job to the resource management system by the GRMS-client. During this process, the locking component takes care of the consistency of files/variables.

An R programming language interface that supports the access to the ACGT services is provided in the ACGT environment. This means that R users and developers will have access to distributed resources in a transparent fashion, as if those resources were local. The complexity of the grid is thus hidden from the user.

Again, accessing the ACGT grid environment requires no changes in the core R implementation. In practice grid access is performed through the call of predefined R functions loaded from a package. R users can make use of the grid technology in a transparent way by passing the functions to be executed in the grid as input to one of those predefined functions (grid.apply) in their local code.

The GridR Session User Interface is structured into 3 different areas:

- **Top Menu** - five buttons: *Connect*, *Logout*, *Create Plot*, *Display History*, *Setup*.
- **Left Area** - 2 areas - *Session Explorer* and *File Explorer*. **Note:** Both areas provide additional actions via a right-mouse-click menu.
- **Right Area** - a window for the console input, a form for entering R commands with two additional buttons and a display area for viewing plots.

The **top menu** area has five buttons you can use to control your GridR Session:

- **Connect** - connects to an ACGT R server and create a GridR Session
- **Logout** - terminates the active GridR Session
- **Create Plot** - creates a file in the working directory containing the latest plot
- **Display History** - displays a history of R commands send to the console input window.

- **Setup** - used to change the ACGT R server (DN name and Host) used for GridR sessions and the file type of the plot files

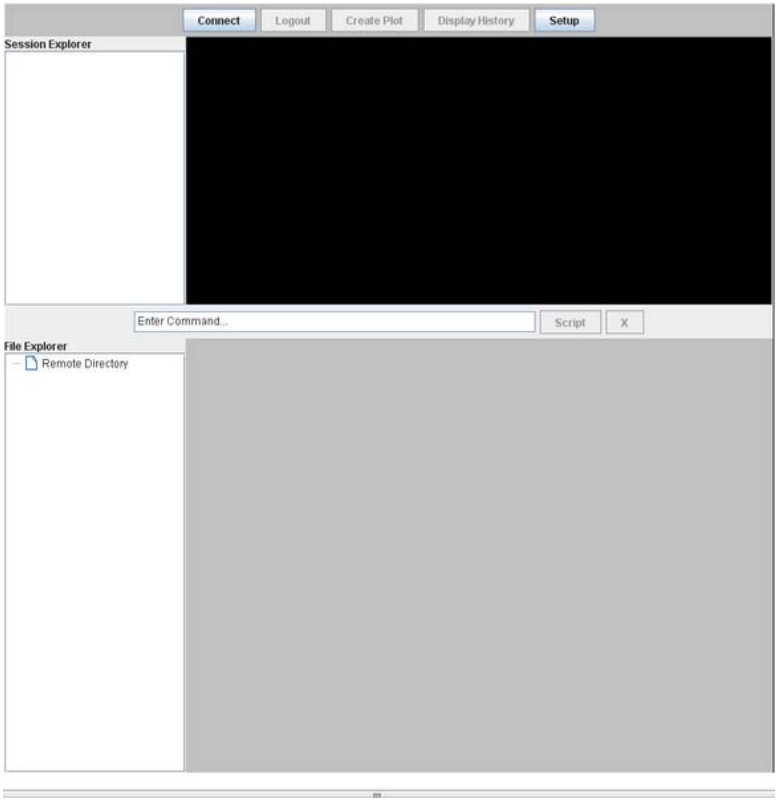

Figure 25 – GridR Session Portlet (P2)

The **left area** shows the multiple GridR sessions that are available:

- The **Session Explorer** can be used to create new sessions, switch between sessions, rename or delete sessions. The active GridR session is shown as "selected". The console input window of this session is shown in the right area of the Grid R tool.
- The **File Explorer** can be used to interact with each session's working directory. Files can be uploaded or downloaded to the local computer from this area. Also, image files containing plots can be previewed in the display area in the right area of the Grid R tool or copied to the clipboard.

The **right area** has three sub-areas:

- **Top-window** - the **Console** area - which shows the active GridR sessions' console input. Any R command or R script entered in the forms below will be shown in this window. This window can be refreshed by sending an empty R command.
- **middle-menu**:
    - an input form marked *Enter Command ...* used for submitting single-line R commands to the console input.

- **Script** button - which allows multi-line R scripts to be sent to the console input window (as a batch of R commands).
- **X** button - which clears the console input.

- **bottom-window** - the **Display** area - which shows previews of plots of an image file selected in the *File Explorer* area.

More information on the GridRSession Portlet can be found in the Handbook: *HB:GridR_Session and HB:Analyzing_data_with_R*.

7.7 Data Access Portlet

In addition to the databases that are connected statically and with a fixed ontology mapping, users of the ACGT environment can also connect SQL databases dynamically to the data mining tools. The **Data Access** Portlet allows users to connect external databases and to execute custom queries on previously added dynamic databases. It was developed by SIVECO with Philips and is available in the *Data Access* area since P2.

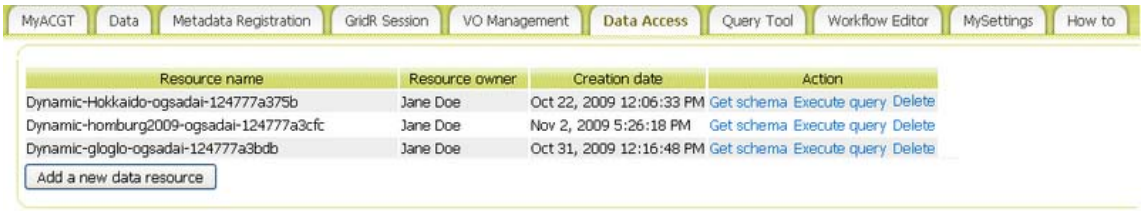

Figure 26 – Data Access Portlet (P2)

The **Data Access** tab lists the previously added databases, with the following information:

- **Resource Name** - Each resource name has the following format: *Dynamic*-USER_SELECTED_NAME-*ogsadai*-UNIQUE_IDENTIFIER.

The *USER_SELECTED_NAME* is what is selected during the import process, and the *UNIQUE_IDENTIFIER* is added automatically by the system

- **Resource Owner** - the user that added the user and can remove it from the ACGT Environment
- **Creation Date**
- **Actions** - simple actions that can be done with the data resource (get schema, execute query, delete)

In prototype 3, all databases added to the ACGT environment are accessible by all users that are members of the *Data Access group*. There is no support for private lists of external databases.

Registering an SQL database can be done by clicking on the *Add a new data resource* button to start the process:

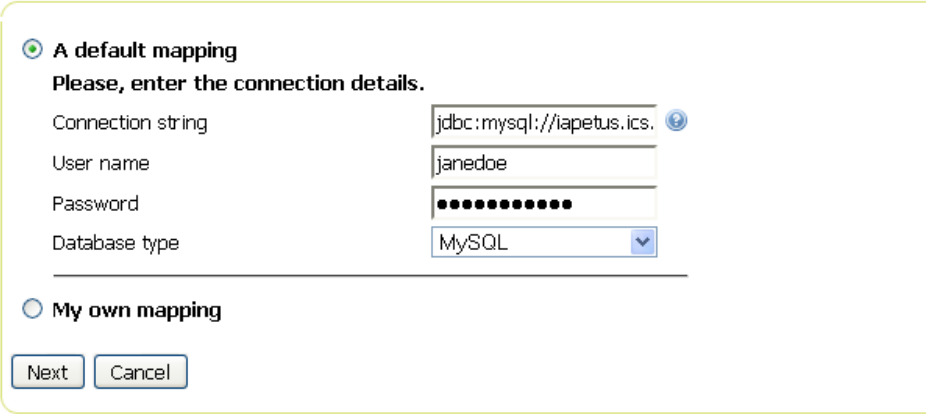

Figure 27 – Data Access Portlet – Adding a new data resource

A database can be added with a default mapping or a custom mapping.

For a **default mapping**, the checkbox corresponding to *A default mapping* must be checked. Then connection details must be provided, including the **location** of the database (the JDBC connection string), the **username** required to access the database, the **password** required to access the database, the **dialect** of the database (*MySQL, PostgreSQL, Oracle*).

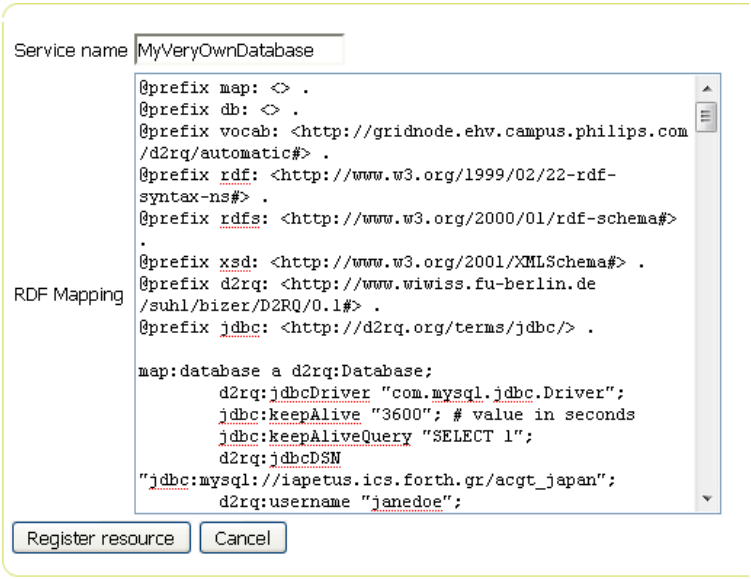

Figure 28 – Data Access Portlet – Default RDF Mapping

After clicking *Next*, a custom name can be provided will be used in the resource's actual name and a default RDF mapping describing the database is shown in an editable form. E.g. a databased named XX, the resource will actually be accessible later with its full name "*Dynamic*-XX-*ogsadai*-nnnnnnnnnn" where an automatic random unique identifier is added to avoid duplicates. Clicking *Register resource* finishes the process.

To add a database with **custom RDF mapping**, the "*My own mapping*" radio button must be checked. A name and RDF mapping must be supplied in the form. The connection details are not requested separately as they are part of the custom RDF mapping.

After adding the data several actions can be performed directly in the Data Access tab or the database can be used in the Workflow Editor.

From the Data Access Portlet, a few simple actions can be done directly in the fourth column of the portlet. The RDF Schema can be retrieved by clicking the *Get Schema* button. The database can be deleted by clicking the *Delete* button.[8]

Also, SPARQL queries can be executed on the data resource by clicking *Execute Query* button. The SPARQL query has to be written or copy-pasted in the form. After reviewing the results of the SPARQL query the results can be saved to the local computer as *.CSV* files (*Comma Separated Values*).

In order to use external data in more complex scenarios, queries on this data resource should be embedded in workflows in the Workflow Editor. A generic SPARQL service must be used to query the external database. The ACGT environment has such a generic service that can be used to add arbitrary SPARQL queries to a workflow - called **Submit SPARQL**.

To use a query on an external database in the Workflow Editor, the SubmitSPARQL service must be added from the left (west) frame of the Workflow Editor. It is available in the tree under *Services* / *Bioinformatics* / *Proxies* / *DynamicDAS* / *SubmitSPARQL.* After dragging it to the workflow (center) area, a box called "SubmitSPARQL" with four input (blue) dots and one output (red) dot appears.

The SubmitSPARQL service must be configured:

- The first blue dot in this box must be connected to an empty space to create the dynamic resource box. Double-clicking it allows users to manually rename it to the name received after registering the dynamic database (ex:*Dynamic-Hokkaido-ogsadai-124777a375b*).
- The second blue dot of the "**SubmitSPARQL**" box must be connected to an empty space to create a SPARQL Query. This box must be double clicked in order to copy-paste or write the SPARQL Query.
- The third (from left to right) blue dot of box (called "DirName") must be dragged to an empty space to create a green box that is the output directory that the service will use to store its results and any temporary files it might need.
- The fourth and last blue dot of the "**SubmitSPARQL**" box must be connected to an empty space. This **FileName** will be used to create a .CSV file where it will store its results.
- The single red dot contains the DMS FileID of the resulting CSV file. This dot can be connected to other components in the workflow (for example an R script using this file as input for analysis).

More information on the Data Access Portlet can be found in the Handbook: *HB:Data_Access and HB:Importing_data*. A step-by-step tutorial is available on the ACGT Portal, named Register and access external databases.

[8] This will only delete the dynamic link and not the data resource itself. The data resource will no longer be available in the ACGT Environment unless it is imported again.

7.8 Query Tool and Mapping Tool Portlets

In the ACGT environment, data from different trials can be accessed *"as if"* it was stored in a single integrated repository. The ACGT environment uses a *master ontology* to represent the domain of clinical trials on cancer and to *semantically integrate* these multiple trial repositories. Using the master ontology, queries can be created on all the various data from clinical trials (provided that the clinical trials are registered). A special language, called SPARQL, is used to create these queries. Due to the complexity of SPARQL, a special user-friendly web-tool called the *ACGT Query Tool* was developed by UPM. The *Query Tool* allows end-users to easily define new queries to access the data from registered trials, hiding the complexity of SPARQL. Another software component, the mediator is used in the ACGT environment to handle queries in terms of the global schema and translate them to sub queries for each of the specific trials.

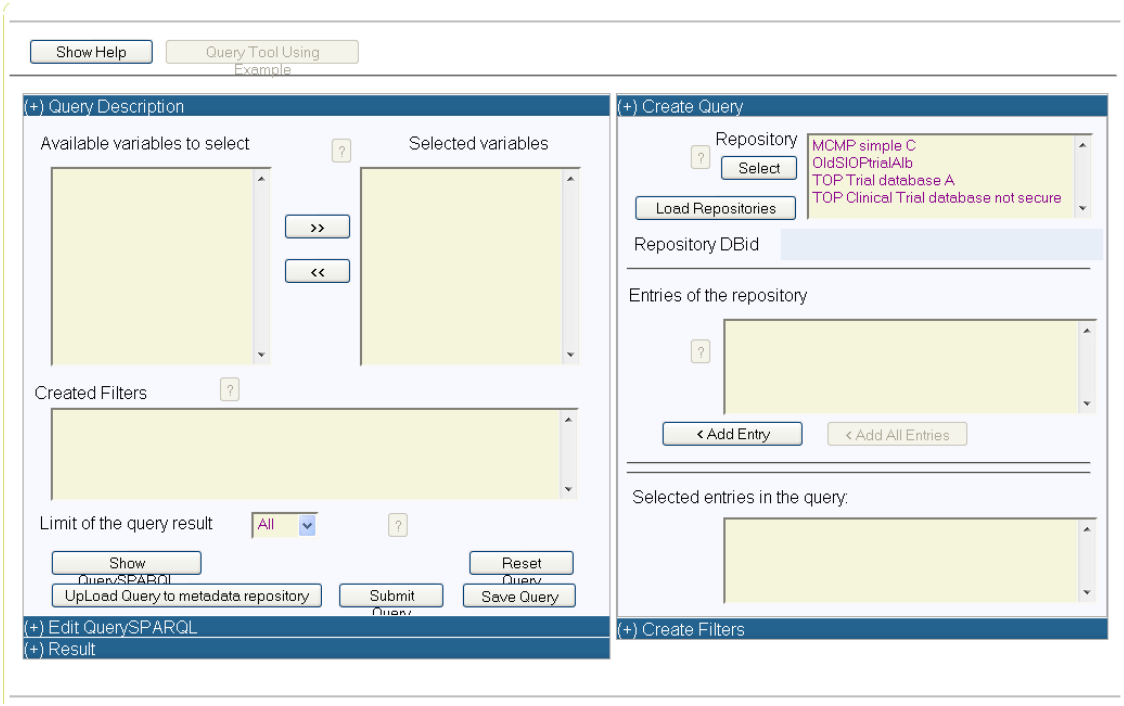

Figure 29 – Query Tool (P3)

The *ACGT Query Tool* can be used retrieve data from clinical registered trials by defining queries and testing them. After testing and refining a query, it can be uploaded to a "query repository" to be used in more complex workflows.

The *Query Tool Interface* is split into two columns:

- the **left column** shows the status of the query so far and allows you to further edit the query (Query Description, editing the resulting SPARQL query, viewing the results of the query)
- the **right column** shows the various available data that can be used as "building blocks" for the query (existing/saved queries, data repositories, descriptions of data from the repositories, filters)

Each column has 3 sections, with a separate title bar (*The default 2 sections, one in each column, are marked with italic*):

- Left column
  - Query Description
  - Edit Query SPARQL
  - Result
- Right column
  - Saved Queries
  - Create Query
  - Create Filters

The following 5 most important areas of the tool are used (typically in this order), to create a query. **Note:** Special numbered help is provided for each section. Pressing the **Show Help** button in the top left of the screen enables all the help buttons.

1. **Create Query** - Repository
2. **Create Query** - Entries of the repository
3. **Query Description** - Variables
4. **Query Description** - Filters
5. **Query Description** - Query Limits

To locate the areas, note that in the list above the first label shows the section title (as it is displayed in its title bar) and the second is the area label which is shown inside the corresponding section. E.g. the *Repository* is shown in the **Create Query** section in the right column.

Step 1. Select the repository

The repository section is located in the right column under the **Create Query** title bar. The registered clinical trial databases that are available to get data from are shown as a list. The ID of the currently selected repository is shown below the list after the *Repository DBid* label. The list of repositories can be refreshed by pressing the **Load Repositories** button. Clicking on a repository in the list and pressing the **Select** button loads the information from the selected repository in the next area.

Step 2. Select entries from the repository

The entries area (*area 2*) is located in the right column under the **Create Query** title bar, below area 1. If a repository is selected in area 1, the list of entries i.e. description of the data that can be retrieved from the repository are shown in the list of *Entries of the repository*. Each entry is a representation of basic information that can be asked about the data in the repository. **Note:** The name of the repository that is currently selected is shown to the right of the *Entries of the repository* label. Two buttons, **Add Entry** and **Add All Entries**, are provided to add the selected or all the entries in the list to be added to the query.

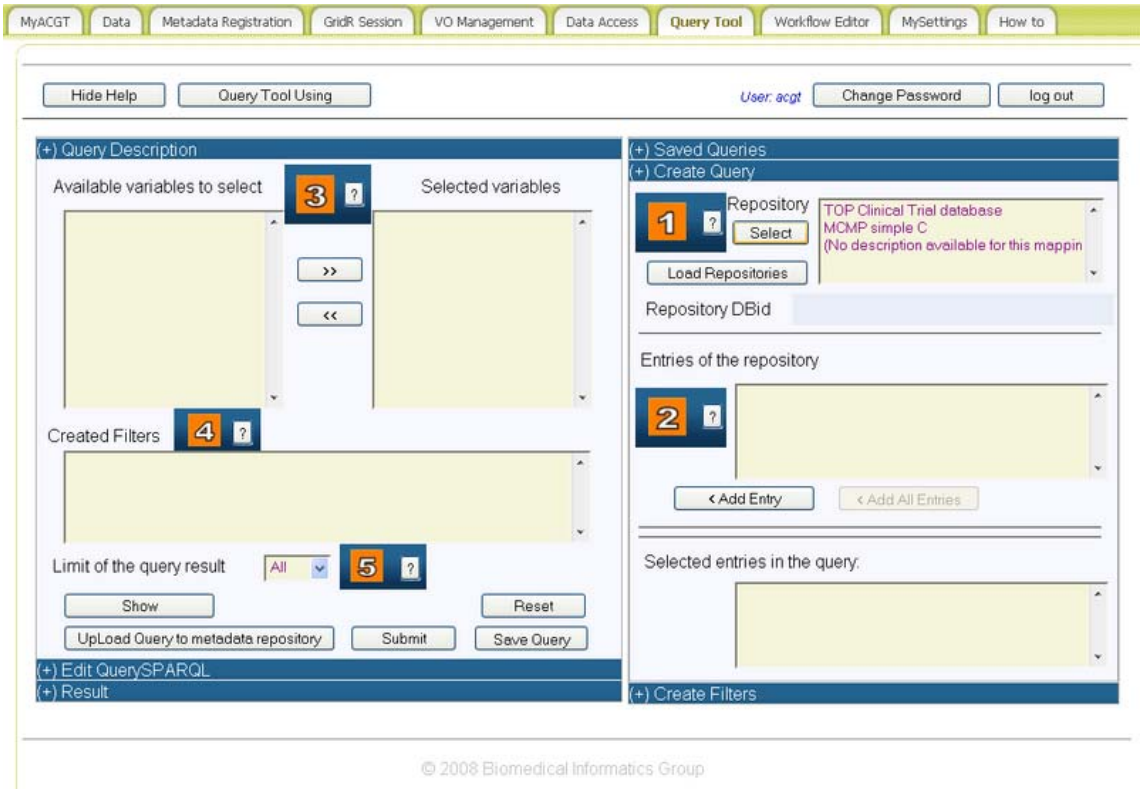

Figure 30 – Query Tool – 5 steps

Step 3. Select variables

The variables area (*area 3*) is located in the left column under the **Query Description**. There are two lists in this area. Two buttons **>>** and **<<** allow the variables to be moved between the two lists.

- The list of **available variables** (*unused*) - each variable is necessary for one or more of the entries selected in area 2
- The list of **selected variables** (*used*)which form the columns of the result of the query

Step 4. Create filters

A filter can be created by selecting an available variable as the *first parameter*, then selecting an operator (**<**,**>**,**=**) and then entering a numerical value as the *second parameter*. After clicking the **Create Filter** button, the filter is added to the created filters list. These filters act as numerical constraints to restrict the results of the query (e.g. selecting only patients with *Age > 30*).

Area 4 actually has two different parts:

- The **Create Filters** section (last section in the right column)
- The **Created Filters** area in the **Query Description** section (in the left column)

Step 5. Limit the results of the query

Area 5 is located in the **Query Description** section (first section, left column) and allows the limitation of the results of the query by specifying the maximum amount of results that a query must generate:

- A drop-down list labeled *Limit of the query result* shows the current maximum limit selected for the results of the query (All, 10, 50, 100 or 500).

Note: The default option is **All** - all possible results will be retrieved after executing the query

The *Query Tool* was developed as a stand-alone web application by UPM and was integrated as an IFrame by SIVECO in the ACGT Portal. The Query Tool has embedded help from its stand-alone version and has a distinct Look & Feel among ACGT Resources and Tools.

More information on the Query Tool can be found in the Handbook: *HB:Query_Tool and HB:Retrieving_data*. A step-by-step tutorial is available on the ACGT Portal, named Querying clinical data.

7.9 Workflow Editor Portlet

The **Workflow Editor** is a web-based tool developed by FORTH allowing users to create, edit, execute and publish workflows in the ACGT workflow environment. The ACGT Workflow Editor Portlet was loosely integrated by SIVECO in the ACGT Portal through an IFrame, as it is a standalone application with a complex user interface, hard to duplicate inside a JSR 168 portlet. The Workflow Editor has embedded help as tooltips.

The **Workflow Editor** aims to assist physicians in their scientific research by supporting the composition of different data access, knowledge extraction and analytical services into **complex workflows**. This way one can extend and enrich the functionality of the ACGT system by reusing *existing* ACGT compliant services and producing "added value" *composite services*.

The **Workflow Editor** and its ACGT workflow environment attempt to bridge the usability gap between grid infrastructure and clinicians. Furthermore, the Workflow Editor takes advantage of the computational and storage power of a grid. This is essential for **scientific workflows** which tend to operate on large amounts of data with complex algorithms. The ACGT workflow environment allows private workspaces (to create and execute your own workflows) and community shared workspace allowing the user community to *exchange information* so that users can benefit from each other's research.

From the **Workflow Editor** a user:

- can view all the available *services*, *R scripts* and *mediator queries* which are located at a remote server in the ACGT Metadata service repository
- can browse and use his/her files from the Grid Data Management Service (DMS) of the ACGT grid
- monitor a workflow which is already running in the enactor.

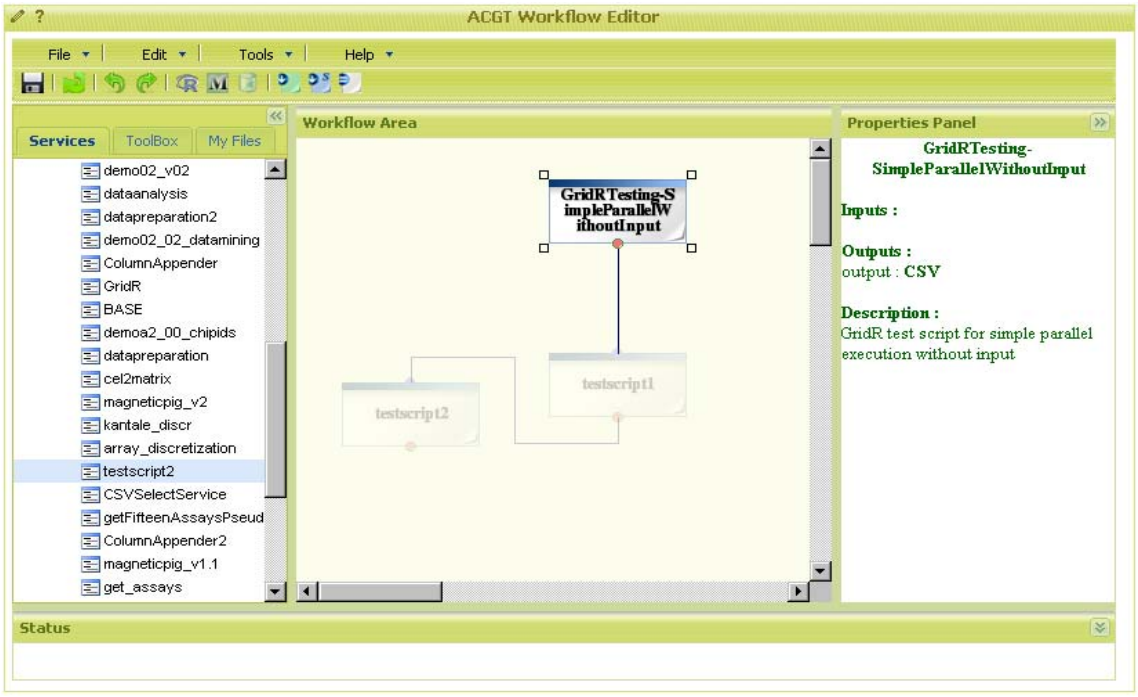

Figure 31 – Workflow Editor Portlet (P2)

Users can:

- *search* and *browse* the available *services* and *data sources*
- create workflows combining available services and data sources through an intuitive and user friendly interface.
- *store* the workflows in their private area and later *retrieve* and *edit* them so that new versions can be produced.
- *publish* and *share* workflows so that other users can use them in their research.
- *execute* the workflows and *monitor* their enactment status.

The user interface of the Workflow Editor consists of five **frames** (areas) (as seen in the figure):

- The "**North**" frame contains a menu bar and menu buttons. Also called **Menu Bar**.
- The "**West**" frame contains all the available services (according to the ACGT web service repository) that can be used to create a workflow and like input/output variables, constants, loop branches. Also called **Service Panel**.
- The "**Center**" frame is the main frame where the workflow is drawn. Also called **Workflow area**.
- The "**East**" frame contains information about the selected (if any) items in the workflow. Also called **Information Panel**.

- The "**South**" frame contains information about the workflow status e.g. response from enactor, errors during deploy/run etc. Also called **Status Panel**.

Figure 32 – Workflow Editor User Interface

More detailed information of the user interface of the Workflow Editor can be found in the ACGT Handbook: HB:Workflow_Editor. A detailed description of what can be done in the Workflow Editor is available in the ACGT Handbook: HB:Setting_up_workflows. A step-by-step tutorial is available on the ACGT Portal: Create and use workflows.

7.10 Service Monitoring Portlet

The service monitoring portlet provides monitoring capabilities for services started by users using the Service Manager Portlet (see D14.2). Its simple design relies on plain JSP for rendering, which gives it robustness. On the other hand the simple approach limits the solutions available for making the user interface dynamic and appealing. This is a common trade-off that has to be addressed when implementing the JSR 168 specifications.

The Service Monitoring Portlet is available in the *MyACGT/ Service Enactings* area since P2. It has become obsolete since it was replaced by the Workflow Editor's monitoring area (bottom pane).

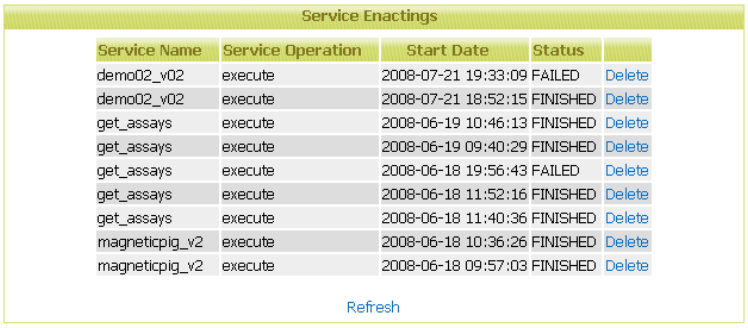

Figure 33 – Service Monitoring Portlet (P2)

7.11 Ontology area

The *Ontology* area allows end-users to find more information about the ACGT Master Ontology, a formal representation of concept and relations in the domain of cancer which is used to structure and describe clinical data. The ACGT Master Ontology is used in the background by ACGT services, so for end-users it may remain invisible. In some situations, however, it may be necessary to have a closer look at the ontology or suggest updates to the ontology.

The *Ontology* area has three subsections:

- *What is the ACGT Master Ontology* – which embeds the ACGT MO page on the ACGT Handbook
- *Ontology Viewer* – embedded as IFrame (External service)
- *Ontology Submission Tool* – embedded as IFrame (External service)

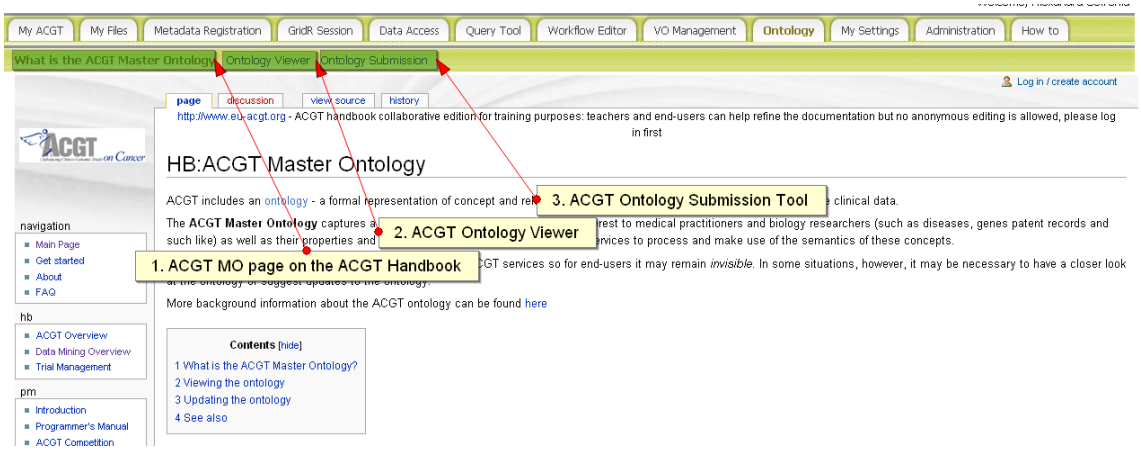

Figure 34 – Ontology Area (P3)

The *ACGT Ontology Viewer* (OV) is a user interface to the Master ACGT Ontology. The Ontology viewer can be opened by clicking on the Ontology Viewer sub tab of the Ontology tab of the ACGT portal. It can be used by registered ACGT users as well as non-registered visitors which access the site directly (although in this case some of the functionalities it

offers are not available for visitors). The Ontology Viewer allows the visualization of the tree-like structure of the ACGT Master Ontology. User-specific views on the ontology can be created. Details on the selected node in the ontology are displayed on the right panel. A search tool allows the user to locate an ontology term in the ontology tree without knowing the hierarchical semantic path leading to it.

More information on the ACGT Ontology Viewer is available in the Handbook http://handbook.eu-acgt.org/HB:Viewing_the_Ontology. A step-by-step tutorial is available on the ACGT Portal.

The *Ontology Submission* section allows users with *Ontology Curation Rights* to login and access the Ontology Management environment. Submission of new terms to the ACGT ontology and ontology curation are only performed by advanced users. Deliverable D7.6 contains a description of the use of these tools. The Ontology Submission Tool is an external service which does not use the ACGT Security infrastructure. More information on the ACGT Ontology Submission Tool is also available in the Handbook http://handbook.eu-acgt.org/HB:Updating_the_Ontology

7.12 How to

The *How to* area contains a brief overview of ACGT, the introduction to the ACGT Handbook as the main information repository, and a list of the latest ACGT tutorials. This area is covered in section *3.4 Accessing help and tutorial in the ACGT Portal.*

8 ACGT Back-end Portlets

8.1 GAS Portlet

Gridge Authorization Service (GAS) is an authorization system, which can be the standard decision point for all components of a system. Security policies for all system components are stored in GAS. Using these policies, GAS can return an authorization decision upon the client request. GAS has been designed such that it is easy to perform integration with external components and it is easy to manage security policies for complex systems. Possibility to integrate with many Globus Toolkit and operating system components makes GAS an attractive solution for grid applications.

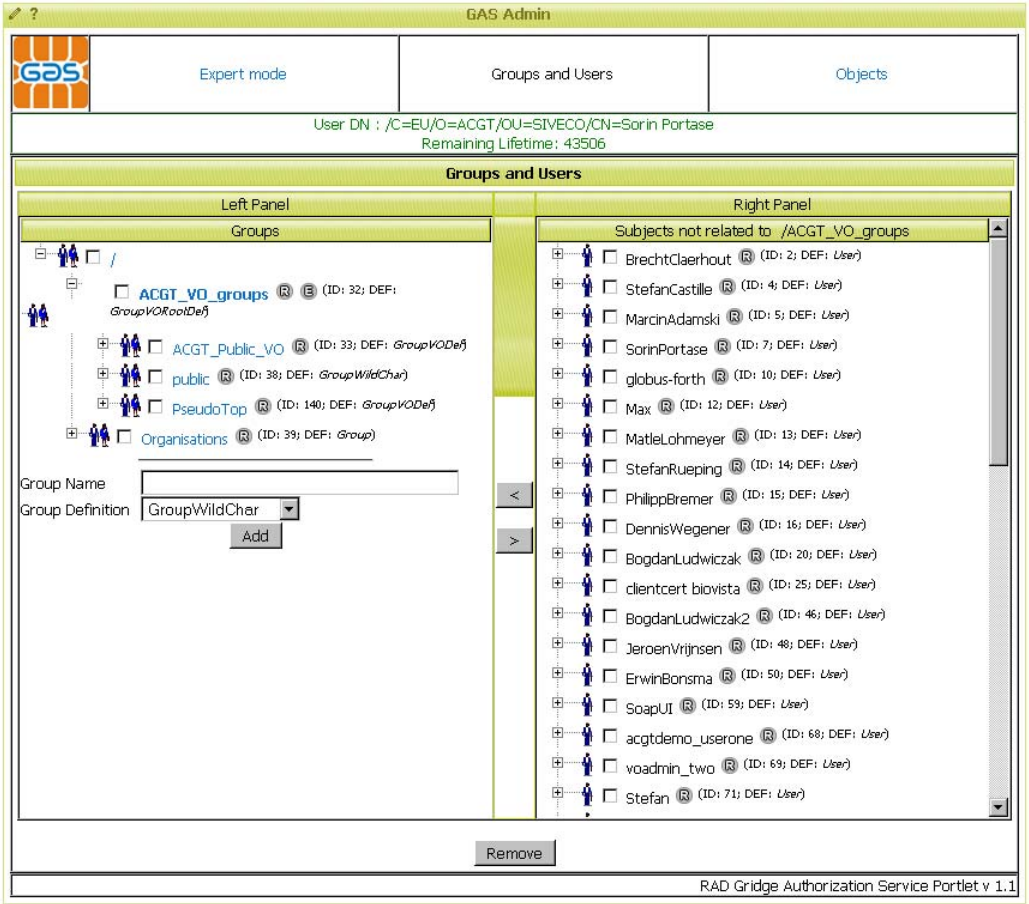

Figure 35 – GAS Portlet (P2)

The main goal of GAS is to provide functionality that would be able to fulfill most authorization requirements of grid computing environments. GAS is designed as a trusted single logical point for defining security policy for complex grid infrastructures. As the flexibility is a key requirement, it is to be able to implement various security scenarios, based on push or pull models, simultaneously.

Secondly, GAS is considered as independent of specific technologies used at lower layers, and it should be fully useable in environments based on grid toolkits as well as other toolkits. The high level of flexibility is achieved mainly through the modular design of GAS and usage of the complex data structure, which can model many scenarios and objects from the real world.

The GAS portlet is an administration portlet application for GAS developed by PSNC and deployed into the ACGT Portal by SIVECO.

The portlet is available in the *GAS* area of the ACGT Portal since P2. It requires *jsrgasportlet* group membership which is not default.

8.2 GRMS Portlet

The **Gridge Resource Management System** (GRMS) is an open source meta-scheduling system, which allows developers to build and deploy resource management systems for large scale distributed computing infrastructures. The GRMS, based on dynamic resource selection, mapping and advanced scheduling methodology, combined with feedback control architecture, deals with dynamic Grid environment and resource management challenges (e.g. load-balancing among clusters, remote job control or file staging support). Therefore, the main goal of the GRMS is to manage the whole process of remote job submission to various batch queuing systems, clusters or resources. It has been designed as an independent core component for resource management processes, which can take advantage of various low-level Core Services and existing technologies. Finally, the GRMS can be considered as a robust system, which provides abstraction of the complex grid infrastructure as well as a toolbox, which helps to form and adapts to distributing computing environments.

Special features of GRMS include:

- Dynamic resource discovery using different information sources to get static and dynamic parameters of machines, queues or jobs in Grid environment.
- Powerful job description providing wide variety of means to express application and user requirements.
- Framework approach to support different Data Management Systems.
- Complex but flexible workflow job support – precedence constraints based on job status changes.

The GRMS Portlet is a graphical user interface client to the GRMS System. The GRMS Portlet was developed by PSNC and deployed into the ACGT Portal. Using GRMS in the ACGT Portal requires *grms portlet group* membership which is not default.

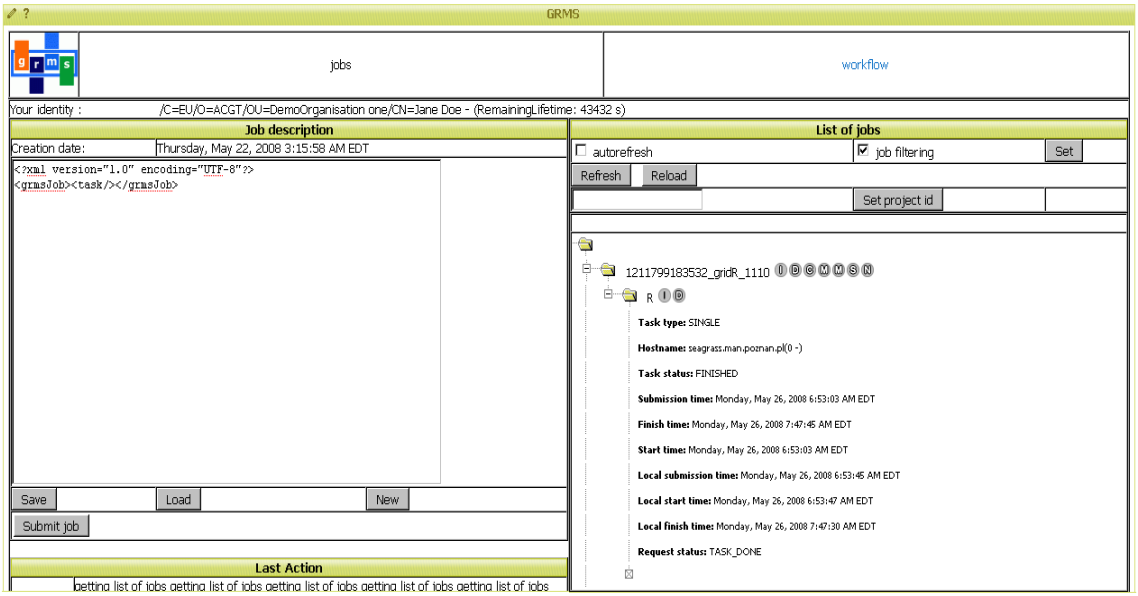

Figure 36 – GRMS Portlet (P2)

8.3 Virtual Organizations Management Portlet

Most of the time, a trial involves many people from different locations that need to access distributed resources. A Virtual Organization (VO) is a group who shares the same computing resources.

The VO Management Portlet is a variation of the GAS Portlet, designed for an easier management of Virtual Organizations (VOs). The VO Management Portlet provides a better approach to a VO approach to roles and rights, while keeping a good control of the resource access. The **VO Management** is a suite of two web-based tools designed to allow end-users to create their own Virtual Organizations (VOs) and to add users and resources to them. The user also has the opportunity to create roles inside their VOs and to assign people and resources to those roles.

The VO Management area of the ACGT Portal contains 4 sub-areas:

- **VO Management** – This application is focused on the management of users belonging to a VO and their roles inside the VO.
- **VO Management Configuration –** Administrator-only configuration settings for *VO Management*
- **VO Objects** - This application is focused on management of the permissions needed to access the VO's resource.
- **VO Objects Configuration** – Administrator-only configuration settings for *VO Objects*

The VO Management Portlet was released in prototype 3 as a back-end proof-of-concept portlet, requiring *vomanagementportlet* group membership which is not default.

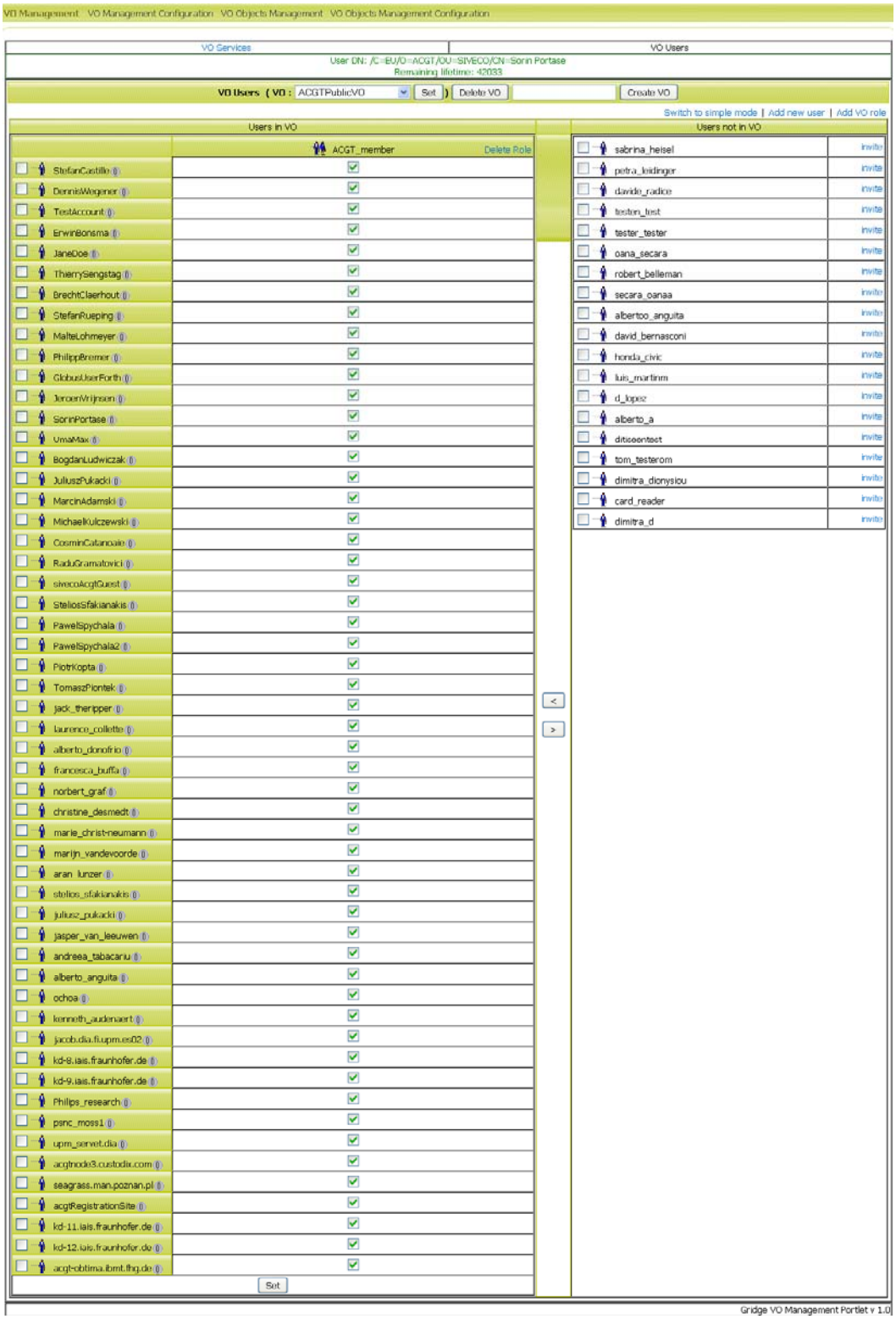

Figure 37 – VO Management Portlet (P2) – VO Users for *ACGTPublicVO*

Further work can be done to transform the *VO Management* and *VO Object* areas to front-end user-friendly areas, while the remaining 2 sub-areas remain as back-end.

9 Browser and network configurations

9.1 ACGT Portal - Browser configurations

The ACGT Portal was developed and tested to ensure cross-browser compatibility: Internet Explorer, Mozilla Firefox and Google Chrome to cover the spectrum of browsers used by potential end-users. Some components, notably the Workflow Editor, do not run on Internet Explorer, due to limitations of the open-source graphical libraries used.

However, the registration and login processes and the GridR Session applet require the Java Runtime Environment (JRE), which might not be installed on client machines. Availability of the Java Runtime Environment version can be checked online at http://www.javatester.org/version.html. The ACGT Portal requires JRE version 1.6+. If JRE is not installed or the installed version is lower than 1.6, the latest version of Java Runtime Environment can be downloaded from http://java.sun.com/javase/downloads/.

Also, the registration and login processes require trusting the ACGT Certification Authority (CA) in order to allow it access to the resources of the client computer in order to create the ACGT Passport.

Trust can be acknowledged by a browser either by:

1. Allowing applets to run when a confirmation window is shown by the browser
2. Adding the following two certificates to the browser's trusted certification authorities:
   - https://acgt.custodix.com/certificates/acgtca.crt
   - https://acgt.custodix.com/certificates/ca.cer

Step-by-step guides for solution 2 for Internet Explorer, Mozilla Firefox and Google Chrome are available in the ACGT Handbook: Trusting_the_ACGT_Certificate_Authorities

Solution 2 is recommended since many services are available using HTTPS protocol (e.g. ACGT Portal, Workflow Editor) using certificates signed by ACGT CA. Therefore adding the ACGT CA to the list of trusted CAs will automatically ensure that HTTPS sites in ACGT will be allowed access without raising a confirmation box for each access.

9.2 ACGT Portal – Network configurations

If a component of ACGT does not appear at all or a generic error message such as *"Connection Error"* appears, there might be a connection problem to that particular server.

A local administrator needs to be contacted, to check if there are firewall rules that prevent the client machine from connecting to ACGT servers and if affirmative, to enable special rules for that ACGT server.

A list of servers and ports that are part of the ACGT Environment:

Server	Port	Description of the component
acgt.siveco.ro	80	ACGT Portal - Login
epimetheus.ics.forth.gr	8080	Alternative ACGT Portal mirror
myproxy.custodix.com	7512	Used in the login process
kd-8.iais.fraunhofer.de	8444	Used by GridR console
moss1.man.poznan.pl	21	Used for file upload and download
ochoa.dia.fi.upm.es	8080	Query Tool
gas.custodix.com	12355	Access to GAS server