Functional & technical specification of the ACGT portal

Project Number: FP6-2005-IST-026996

Deliverable id: D 14.1

Deliverable name: Functional & technical specification of the ACGT portal

Date: 14 September, 2006

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 14.1
Document name:	Functional & technical specification of the ACGT portal
Document type (PU, INT, RE)	RE
Version:	2.0
Date:	14.9.2006
Editor: Organisation: Address:	Radu Gramatovici SIVECO 8-10 Bd. Maresal Averescu, Bucharest, Romania

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

ABSTRACT:

The present deliverable presents the ACGT Portal. The ACGT Portal is intented to represent to main access gate to the ACGT grid platform. The document is based on the "User requirements and specification of the ACGT internal clinical trial" document, published as deliverable D 2.1. The present document consolidates the requirements specified in D 2.1, form the point of view of the ACGT Portal.

Part 1 of this document presents the state of the art in the area of biomedical portals and more general, of Web portals. In **Part 2**, some main portal user scenarios are described. **Part 3** of the document presents the skeleton of the ACGT Portal functionalities. **Part 4** describes technical specifications of the ACGT Portal.

KEYWORD LIST: Web Portal, Grid Portal, Biomedical Portal

MODIFICATION CONTROL			
Version	Date	Status	Author
0.1	20.7.06	Draft	Radu Gramatovici
0.9	30.8.06	Draft	Radu Gramatovici
1.4	14.9.06	Draft	Radu Gramatovici
2.0	16.11.06	Final	Radu Gramatovici

List of Contributors

- Florin Manea, SIVECO Romania
- Dawid Szejnfeld, PSNC
- Otto Zelch, SIVECO Romania

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). Biomedical data and information to be considered include not only clinical information relating to tissues, organs or personal health-related information but also information at the level of molecules and cells, such as that acquired from genomics and proteomics research.

ACGT focuses on the domain of Cancer research, and its ultimate objective is the design, development and validation of an integrated Grid enabled technological platform in support of post-genomic, multi-centric Clinical Trials on Cancer. The driving motivation behind the project is our committed belief that the breadth and depth of information already available in the research community at large, present an enormous opportunity for improving our ability to reduce mortality from cancer, improve therapies and meet the demanding individualization of care needs.

In addition to the shear volume, data collected using a variety of laboratory technologies and techniques are often published without the background information (method of capture, sample preparation, statistical techniques applied) that is needed to reproduce results. In fact, a typical researcher spends as much time trying to understand the origins of a dataset as actually performing new analyses. Rarely is a clinical biostatistician able to make good use of data collected on studies in which they were not directly involved with, largely due to incomplete or non-existent annotation and standardization of the information. Even within a single laboratory, researchers have difficulty integrating data from different technologies because of a lack of common standards and other technological and medico-legal and ethical issues. As a result, very few cross-site studies and clinical trials are performed and in most cases it isn't possible to seamlessly integrate multi-level data. Moreover, apart from problems in sharing and re-using data, what is even more critical is the fact that clinical researchers or molecular biologists often find it hard to exploit each other's expertise due to the absence of a cooperative environment which enables the sharing of resources and tools for comparing results and experiments, and a uniform platform supporting the seamless integration and analysis of disease-related data at all levels.

The ultimate objective, therefore, of the ACGT project is the development of European Knowledge Grid infrastructure offering high-level tools and techniques for the distributed mining and extraction of knowledge from data repositories available on the Grid, leveraging semantic descriptions of components and data and offering knowledge discovery services in the domain of Cancer research. Special emphasis will be given to the trust that needs to be embedded in the platform and relevant ethical issues, thus creating optimal conditions for service uptake.

In the architecture of the ACGT Grid infrastructure, the portal is the gate to the services. Structured on several layers of access rights, the ACGT Portal is offering information to all the actors playing in the ACGT project and more general in the domain of Cancer research.

The document is based on the "User requirements and specification of the ACGT internal clinical trial" document, published as deliverable D 2.1. The present document consolidates the requirements specified in D 2.1, form the point of view of the ACGT Portal.

For the self-containement of the document, some relevant parts of the D 2.1 document are integrated here (as the introduction presenting general information on the project and the state-of-the-art of Web portals). Other parts of the D 2.1 document (as some parts from the ACGT methodology for user requirements engineering and the classification of the ACGT stakeholders and users) are developed from the perspective of the ACGT Portal.

The present document is structured in four parts.

Part 1 of this document presents a short introduction to the ACGT project, containing the project background, the ACGT environment, the project vision and specific objectives. This short introduction is followed by a review of some methods and principles enumerated in the ACGT methodology for user requirements engineering, interpreted and consolidated from the perspective of the ACGT Portal. These principles that are extremely relevant for the development of the ACGT Portal are the employment of scenarios and prototypes as well as the volatility of requirements through different stages of the portal development. Part 1 ends by a summary of the state-of-the-art in the area of biomedical portals and more general, of Web portals.

Part 2 of the document describes the users of the ACGT Infrastructure, which are also the users of the ACGT Portal. The users are classified through several criteria, like:

- Stakeholder relevance
- End-users
- Data access.

Using the consolidated information on the types of users, the characterization of 10 roles in the ACGT Portal is given.

These roles are:

- Administrator
- Editor
- Reviewer
- Cancer organization representative
- Researcher
- Clinician
- Auditor (member of a regulatory agency)
- Technology supplier
- Standard bodies representative
- Patient

For each of these roles, usage scenarios are described using the basis of privilegies (each role is a collection of privilegies).

Part 3 of the document presents the main ACGT Portal functionalities. The features of the ACGT Portal are grouped under 7 main functionalities that are considered to be most relevant and to bring the most value to the ACGT Infrastructure.

These main features are:

- User personalization
- Search methods
- User activity repository
- User online assistance
- Detailed scheme of privilegies and roles
- Data and resource security
- Collaborative work

Each of these features is decomposed in subfeatures and described in detail.

Part 4 describes technical specifications of the Gridsphere platform, which will serve as basis for the development of the ACGT Portal.

The main reasons for choosing to build up the ACGT Portal on the Gridsphere Portal are:

- Standards compliance
- Open-source platform
- References in the domain of biomedical portals.

PART 1

State of the art in Biomedical Web Portals

1 Introduction

1.1 Project background

Recent and forthcoming developments in genomics and the increased importance of genetics in healthcare are already changing clinical care. Research on the molecular mechanisms of cell growth, apoptosis and differentiation has resulted in a better understanding of the nature of cancer cells. The genotypic knowledge of a cancer cell helps to identify the predisposition of the disease and develops therapies adapted to the genotype of a cancer patient. Medicine is getting more individualised.

The information from genetic and protein studies, clinical trials, and other research is growing rapidly. On the other hand there is no unifying infrastructure or common standard for the technologies that cancer researchers use. There are for example no mechanisms for easily sharing and joining data. In responding to these challenges, Biomedical Informatics is quickly evolving into a research field that encompasses the use of all kinds of biomedical information, from genetic and proteomic data to image and clinical data associated with various levels of the human body. This kind of integration and exploitation of the data and information requires a new synergetic approach that enables a bi-directional dialogue between these scientific disciplines and integration in terms of data, methods, technologies, tools and applications. While the goal is clear, the path is difficult to go, fraught with technical, scientific, clinical, legal and ethical challenges. Many new tools for today's biomedical researcher have been developed to find the mechanism behind cancer, whereas legal and ethical issues are lagging behind.

The main objective of ACGT is the fight against cancer. To achieve this goal ACGT has the following objectives

- The ACGT project sees its mission to develop a GRID platform to support and stimulate further exchanges of both clinic and genetic information.
- ACGT intends to trigger the emergence of latent ***clinico-genomic synergies*** to ensure ***faster diagnosis*** and more ***efficient therapy***
- ACGT targets two major cancer diseases namely, ***breast cancer*** *(BRCA)* and ***paediatric nephroblastoma*** *(PN)* presented by three (running) clinical trials.
- In addition, ***in-silico*** oncology trial scenarios will be run to assess the utility of ***tumour-growth simulation*** on both BRCA and PN.

1.2 The ACGT Environment

ACGT was set up to respond to the challenges arising from three global factors as mentioned above:

- Changing environment comprising a number of issues in all areas of life science
- Changes in healthcare delivery comprising the move towards a more individualised medicine and
- Technology push in conjunction with Biomedical Informatics

ACGT focuses on clinical trials on Cancer (Wilms tumor, Breast) and is based on the principles of open access (among trusted partners) developing open source products.

ACGT will provide a unified technological infrastructure to facilitate

- integrated access to multi-level biomedical data;
- The development or re-use of open source analytical tools, accompanied with the appropriate meta-data allowing their discovery and orchestration into complex workflows.

ACGT brings together internationally recognised leaders in their respective fields, with the aim to deliver to the cancer research community an integrated clinicogenomic ICT environment enabled by a powerful GRID infrastructure. In achieving this objective ACGT has formulated a coherent, integrated workplan for the design, development, integration and validation of all technologically challenging areas of work. Namely:

- Grid: delivery of a European Biomedical Grid infrastructure offering seamless mediation services for sharing data and data-processing methods and tools, and advanced security;
- Integration: semantic, ontology based integration of clinical and genomic/proteomic data - taking into account standard clinical and genomic ontologies and metadata;
- Knowledge Discovery: delivery of data-mining GRID services in order to support and improve complex knowledge discovery processes;
- Legal and ethical issues: development and integration of technical solutions regarding data protection and secure personal data management in a European context.

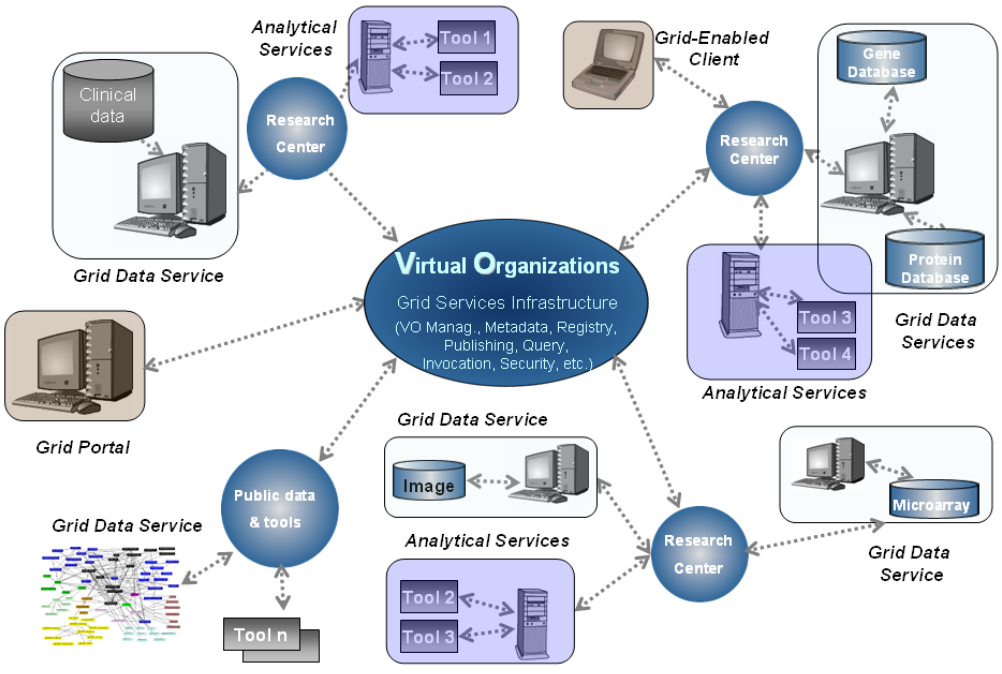

Figure 1: The ACGT Virtual Organizations

The technological platform of ACGT will be validated in concrete setting of advanced clinical trials on Cancer. Pilot trials have been selected based on the presence of clear research objectives, raising the need to integrate data at all levels of the human being.

1.3 Vision of the Project

Information arising from post-genomics research, and combined genetic and clinical trials on one hand, and advances from high-performance computing and informatics on the other is rapidly providing the medical and scientific community with new insights, answers and capabilities. The breadth and depth of information already available in the research community at large, present an enormous opportunity for improving our ability to reduce mortality from cancer, improve therapies and meet the demanding individualization of care needs.

⇨ *A critical set of challenges, however, currently inhibit our capacity to harvest these opportunities.*

Capitalizing on the opportunities apparent to cancer research community requires the integration and exploitation of data and information generated at all levels (from molecular to organ and disease to the population) by the disciplines of bioinformatics and medical informatics, including medical imaging. So, the new raising biomedical informatics (BMI) technology will be utilised and accordingly enhanced in order to meet the posted needs and technological challenges.

This, in turn, requires a new synergetic approach that enables a bi-directional collaboration among these scientific disciplines and integration in terms of data, methods, technologies, tools and applications. In the new area of "genomic medicine", designers and analysts of clinical trials must consider various issues, such as:

- how to design experiments for obtaining coherent and consistent medical and biological data, while avoiding various types of biases and errors,
- how to develop methods for heterogeneous (e.g., genomic, medical) data source integration, including the use of ontologies which facilitate mapping and information retrieval,
- how to develop methods for data selection, checking, cleaning, and pre-processing of combined genomic/medical data, and
- How to incorporate collaborative approaches to data analysis, since biomedical statisticians and data miners in genomics and medicine have been following different methodologies, and dedicated, often proprietary, tools.

In order to address all the previously stated issues, there is an obvious need for an integrated and high-performing Grid-enabled Clinico-Genomic Environment that will not only allow the ubiquitous access to computational resources but it will also enhance the cross-organizational resource sharing. The ultimate objective therefore of the ACGT project is the development of European Knowledge Grid infrastructure offering high-level tools and techniques for the distributed mining and extraction of knowledge from data repositories available on the Grid, leveraging semantic descriptions of components and data and offering knowledge discovery services in the domain of biomedical informatics. Special emphasis will be given to the trust that needs to be embedded in the platform and relevant ethical issues, thus creating optimal conditions for service uptake. Most importantly, such an environment will facilitate the design, monitoring, and evaluation of clinical trials and experimental treatment protocols.

1.4 The ACGT specific objectives

In order to achieve its vision and goals, the project will create and test an infrastructure for cancer research by using a:

> *virtual web of trusted and interconnected organizations and individuals to leverage the combined strengths of cancer centers and investigators and enable the sharing of biomedical cancer-related data and research tools in a way that the common needs of interdisciplinary research are met and tackled.*

A major part of the project is devoted to research and development in infrastructure components that eventually will be integrated into a workable demonstration platform upon which the selected use cases (the Clinical Pilots) can be demonstrated and evaluated against user requirements defined at the onset of the project.

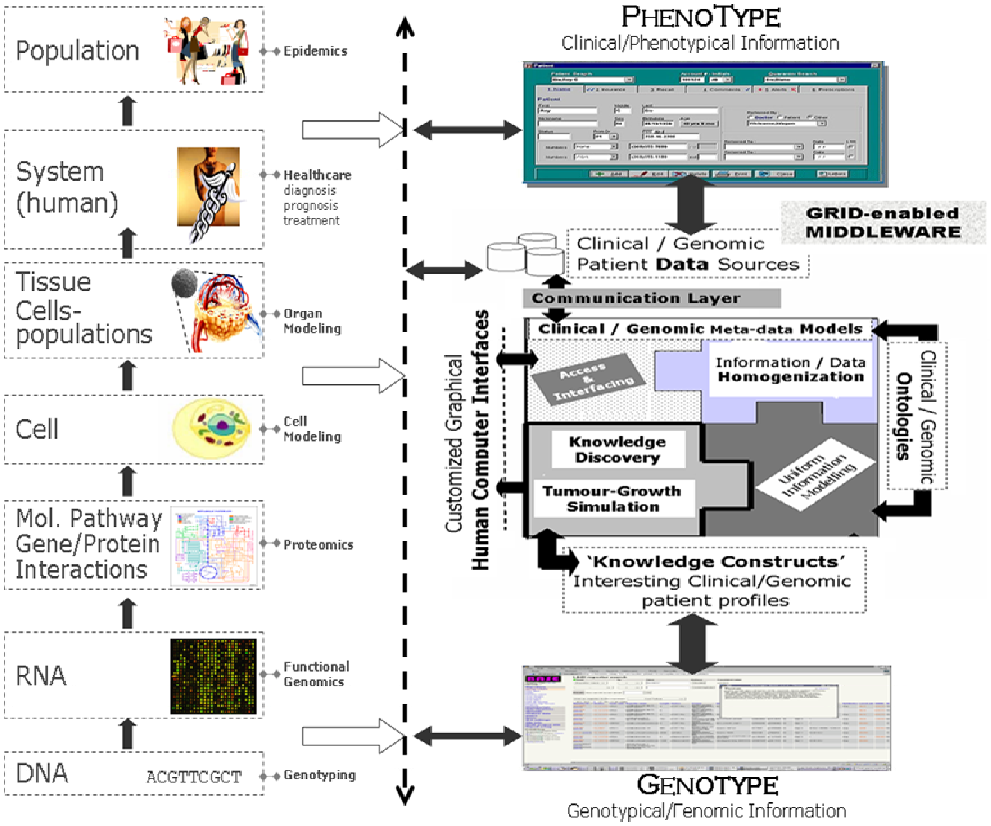

Figure 2: The envisioned ACGT Grid-enabled infrastructure and integrated environment – integration to be achieved at all levels, from the molecular to system and to the population.

In Figure 2 (previous page), an outline schema of the foreseen *ACGT* infrastructure and integrated clinico-genomic environment is shown. Integration targets all levels – from molecular to the human and the population. Grid-enabled mediation functionality (realised by respective services) enable knowledge-enriched, effective and reliable integration.

The principal ***user service objectives*** of the ACGT project can be summarized as:

- Secure that the services reflects the identified stakeholder priorities and validate them in reference to user priorities, i.e. functionality, security and user acceptance.

- Validate the economic feasibility of sustainability of the platform and elaboration of alternative exploitation models, bearing in mind the open source-open access principles on which the project was build.
- Assess the potential usage of the ACGT platform in other domains.

The principal ***business objectives*** of the ACGT project can be summarized as:

- Develop suitable business models based on value creation and value nets.
- Contribute to the creation of a "culture" of sharing of biomedical data and tools based upon common standards and their utilisation in a manner that protects data privacy and security.
- Build a critical mass for the commercial exploitation of ACGT.

The principal ***technological objectives*** of the ACGT project can be summarized as:

- Integration of Clinical Research Centers on Cancer with varying needs and capabilities in a common network for sharing data, applications, and technologies.
- Development of a useable and scalable biomedical Grid that Clinical Research Centers on Cancer will actively use for added value clinical trials.
- Development of new open source tools for multilevel, biomedical data analysis and knowledge discovery as well as adaptation/modification of existing ones so that they comply with the ACGT integration guidelines and architecture, utilise the advantages of Grid computing, and enable high-performing data-mining and biomedical knowledge extraction operations.
- Development of a Master Ontology for Clinical Trials and Cancer
- Enable ontology-based implementation of key Clinical Trial Management Modules.
- Utilisation of clinical trial management systems based on standards-based and components-based clinical trial management systems, integrative cancer research applications and innovative tools to support (a) ontology-based integration and sharing of data and biomedical information and (b) advanced data mining and biomedical knowledge extraction.
- Sharing of biomedical information and data upon common standards and utilisation of and in a manner that protects data privacy and security.
- Fostering common usage of vocabularies, common data elements and the formation of a unifying architecture for the support of the advanced clinico-genomics clinical trials of the future.
- Validate the technical performance of the platform to ensure that it fully supports the identified business needs at acceptable performance.

In achieving these objectives the main technological challenges of the project are:

- **Semantic Grid** services, enabling large-scale (semantic, structural, and syntactic) interoperation among biomedical resources and services;
- Master **ontology** (on Cancer) through semantic modelling of biomedical concepts using existing ontologies and ontologies developed for the needs of the project;
- Open source bioinformatics tools and other **analytical services**;
- Semantic **annotation** and **advertisement** of biomedical resources, to allow **metadata-based discovery** and query of tools, and services;

- Orchestration of data access and analytical services into **complex eScience workflows** for post genomic clinical research and trials on cancer;
- **Meta-data descriptions of clinical trials** to provide adequate provenance information for future re-use, comparison, and integration of results;

1.5 Purpose and Structure of this Document

The present deliverable is the D14.1: "Functional & technical specification of the ACGT portal". This document represents a specific consolidation of the user needs and requirements analysis published as deliverable D2.1.

The purpose of the work done and reported here is to:

- Interpret and consolidate the user requirements with respect to the ACGT Portal.
- Define the ACGT Portal in terms of functional and technical specifications.

The present deliverable is articulated as follows:

Part 1 of this document presents a short introduction in the ACGT project containing the project background, the ACGT environment, the project vision and specific objectives. This short introduction is followed by a review of some methods and principles enumerated in the ACGT methodology for user requirements engineering interpreted and consolidated from the perspective of the ACGT Portal. These principles that are extremely relevant for the development of the ACGT Portal are the employment of scenarios and prototypes as well as the volatility of requirements through different stages of the portal development. Part 1 ends by a summary of the state-of-the-art in the area of biomedical portals and more general, of Web portals.

Part 2 of the document describes the users of the ACGT Infrastructure, which are the also the users of the ACGT Portal. The users are classified through several criteria, like:

- Stakeholder relevance
- End-users
- Data access.

Using the consolidated information on the types of users, the characterization of 10 roles in the ACGT Portal is given:

- Administrator
- Editor
- Reviewer
- Cancer organization representative
- Researcher
- Clinician

- Auditor (member of a regulatory agency)
- Technology supplier
- Standard bodies representative
- Patient

For each of these roles, usage scenarios are described using the basis of privilegies (each role is a collection of privilegies).

Part 3 of the document presents the main ACGT Portal functionalities. The features of the ACGT Portal are grouped under 7 main functionalities that are considered to be most relevant and to bring the most value to the ACGT Infrastructure.

These main features are:

- User personalization
- Search methods
- User activity repository
- User online assistance
- Detailed scheme of privilegies and roles
- Data and resource security
- Collaborative work

Each of these features is decomposed in subfeatures and described in detail.

Part 4 describes technical specifications of the Gridsphere platform, which will serve as basis for the development of the ACGT Portal.

The main reasons for choosing to build up the ACGT Portal on the Gridsphere Portal are:

- Standards compliance
- Open-source platform

References in the domain of biomedical portals.

2 The ACGT methodology for the Engineering of User Requirements

In D2.1 “User requirements and specification of the ACGT internal clinical trial”, a special section is devoted to the ACGT methodology for the engineering of user requirements.

Some principles that are of extremely importance for the development of the ACGT Portal and training modules are presented below.

2.1 Specific elicitation techniques

A number of techniques exist and are often used during the requirements elicitation phase of a project. They include:

- Interviews
- Scenarios
- Soft systems methods
- Prototyping
- Observations and social analysis
- Requirements reuse

The complexity of the domain which is addressed by the ACGT project necessitated that a spiral process of requirements analysis, elicitation, documentation and validation is adopted. Specific techniques have also been selected for the elicitation, negotiation and agreement of requirements as well as their validation. These techniques are ***scenarios*** and ***prototyping***.

2.1.1 Scenarios

"Scenarios are arguably the starting point for all modelling and design" [THO1992]. They allow us to take a backward glance. They use a simple, traditional activity – storytelling – to provide a vital missing element, namely a view of the whole of a situation. 'A straight-line sequence of steps taken by independent agents playing system roles' is roughly what most engineers mean by Scenario. Synonyms include Operational Scenario – itself part of a Concept of Operations, (Test) Case (of actions or events). A sequence of numbered steps in the form '<role> does <action>' is a simple and effective way to describe how a result is to be obtained.

Scenarios are a powerful antidote to the complexity of system development [ROL1998]. 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 analyses, but are almost always based on the idea of a sequence of actions carried out by intelligent agents. People are very good at reasoning from even quite terse stories, detecting inconsistencies, omissions, and threats with little effort. These innate human capabilities give scenarios their power. Scenarios are applicable to systems of all types, and may be used at any stage of the development life-cycle for different purposes [SUT2003].

Scenarios are stories which explain how a system might be used. They should usually include:

- a description of the system state before entering the scenario;
- the normal flow of events in the scenario;
- exceptions to the normal flow of events;
- information about concurrent activities;
- a description of the system state at the end of the scenario.

2.1.2 Prototypes

A prototype is an initial version of a system which may be used for experimentation. Prototypes are valuable for requirements elicitation because users can experiment with the system and point out its strengths and weaknesses. The prototype allows users to experiment and discover what they really need to support their work, but more importantly forces a detailed study of the requirements which reveals inconsistencies and omissions.

Rapid development of prototypes is essential so that they are available early in the elicitation process. Establishes feasibility and usefulness before high development costs are incurred.

Prototypes are also essential for developing the 'look and feel' of a user interface and they can be used for system testing and the development of documentation.

A prototype of a proposed system is presented to workers for critical comments. Revisions are made to the original prototype, producing a second version that is again presented to users for critical analysis. The process of revising and submitting to users continues until some criterion for acceptability is reached. [THO1992] advocates the use of rapid prototyping as a vehicle for allowing technical communicators to become a part of the development team.

2.1.3 Volatility of Requirements

Requirements change. During the time it takes to develop a system the users' needs may mature because of increased knowledge brought on by the development activities, or they may shift to a new set of needs because of unforeseen organizational or environmental pressures.

If such changes are not accommodated, the original requirements set will become incomplete, inconsistent with the new situation, and potentially unusable because they capture information that has since become obsolete. One primary cause of requirements volatility is that "user needs evolve over time" [SAG1990]. The requirements engineering process of elicit, specify, and validate should not be executed only once during system development, but rather should be returned to so that the requirements can reflect the new knowledge gained during specification, validation, and subsequent activities. A requirements engineering methodology should be iterative in nature, "so that solutions can be reworked in the light of increased knowledge" [SUD2005].

Another cause of requirements volatility is that the requirements are the product of the contributions of many individuals, and these individuals often have conflicting needs and goals. For example, there usually is more than one stakeholder, with each one having different and often contradictory views and interests. Due to political climate and other factors, the needs of a particular group may be overemphasized in the elicitation of

requirements. Later prioritization of the elicitation communities' needs may correct this oversight and result in requirements changes. Both the traceability of requirements and their consistency may be affected if these changes are frequent and not anticipated.

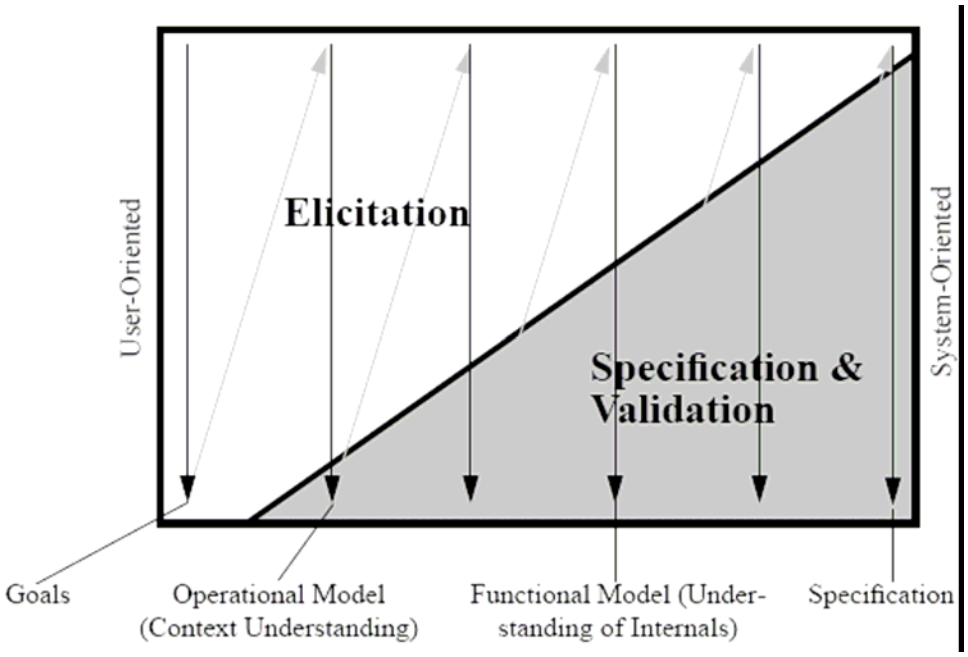

Figure 3: Requirements Engineering as an Iterative Process

Organizational complexity is another cause of requirements volatility. Organizational goals, policies, structures, and work roles of intended end users all may change during the course of a system's development, especially as the number of users affected by a system's development increases. An iterative process of requirements development can address the problems of volatility [SUT2002]:

> The traditional notion of the software development life-cycle with requirements capture being completed before the design stage is no longer satisfactory. Requirements capture and design are now seen to be symbiotic. The initial set of requirements needed to start off the design process is gradually refined into a systematic and coherent statement of requirements hand in hand with the refinement of design.

Due to the problems of understanding and scope, user needs may not be clearly expressed initially in the requirements, and the developer or requirements analyst may make some incorrect assumptions based on this ambiguity. With an iterative process (see Fig. 5), those mistaken assumptions can be detected faster and corrected sooner.

2.2 Relevance for the development of the ACGT Portal

The ACGT Portal and training modules represent the interface of the whole ACGT system in front of the users. This is the reason for which all the technics and facts presented above, as user scenarios, protyping and requirements volatility are extremely applicable in the development of the ACGT Portal.

User **scenarios** presented in Chapter 5 of D2.1 were analysed from the point of view of the ACGT Portal. From these user scenarios reflecting different activities that will be carried on within the ACGT infrastructure were abstracted those relevant components that will be performed through the ACGT Portal. These scenarios, grouped on types of users, are presented in Chapter 5 of this document.

Prototypes are also essential for a smooth and sound development of the ACGT Portal. Prototypes wiil be realized in different phases of the development, both for the Web portal and for the training modules.

Prototypes are extremely important for presenting different stages of the portal development, since:

- On one hand, the ACGT portal should be accessible to the users from its early stages, in order to be adopted by the users as the main access point of the ACGT system.
- On the other hand, it is highly expected that the requirements for the ACGT Portal wiil constantly evolve with the ACGT infrastructure and any prediction and analysis made at this point can be invalidated by the future evolution (see the above Section 2.1.3 **Volatility of Requirements**).

A first visual prototype of the ACGT was designed in the analysis phase and is presented in this document (Chapter 6). The visual prototype is made from several page layouts illustrating the main functionalities that the ACGT Portal aims to provide. These pages contain:

- The login page (**Figure 4:** ACGT Portal – Login page prototype)
- A user's homepage (**Figure 5:** ACGT Portal – User's homepage prototype)
- A user's profile page (**Figure 6:** ACGT Portal – User's profile page prototype)
- The workflows page (**Figure 7:** ACGT Portal – Workflows page prototype)
- A specific workflow's page (**Figure 8:** ACGT Portal – Specific workflow page prototype)

A functional prototype of the ACGT Portal developed on the Gridsphere platform (see Chapter 7) is expected by the month 12 of the project. The functional prototype of the ACGT Portal has to contain the main features of the ACGT Portal as presented in Chapter 6:

The prototype of the ACGT Portal will be constantly improved at least until the ACGT infrastructure will reach a stable state.

The same approrach will be taken regarding the training modules. Training is seen as one of the facilities provided by the ACGT Portal. Task 14.3 of the project planned for the first half of the second year of the project will establish the standards and prototypes of the training modules designed to be integrated in the ACGT Portal.

The right approach is to provide the training as much as possible in an online form accessible permanently to all registered users of the ACGT system. The provision of the training should not be limited in time or in terms of access rights.

The training modules shall be developed according to the actual standards (both technical and pedagogical) in this area. Training modules should be simple to use, relevant and based on creating competencies and transferring knowledge rather than information. Task 14.3 “Training modules for clinical and biological investigators and students” of the project should provide a model for how the training modules have to be developed. Within this task, prototypes of training modules will be developed and presented to the ACGT participants.

This risk can be approached by defining a methodology for training content development as an integrated part of the ACGT infrastructure development. Each service or content provider that uses the ACGT infrastructure should be encouraged to create and provide online training modules for its own services or resources. Task 14.3 should also provide a model of how the training modules have to be integrated in the ACGT system.

2.3 References

[ROL1998] Rolland C., Souveyet C., Ben Achour C., *Guiding Goal Modelling Using Scenarios*. IEEE Transactions on Software Engineering, special issue on Scenario Management, 1998.

[SUT2002] Sutcliffe A., User-Centred Requirements Engineering - Theory and Practice, Springer, 2002, ISBN: 1852335173.

[SUT2003] Sutcliffe A., Scenario-Based Requirements Engineering, p. 320, 11th IEEE International Requirements Engineering Conference (RE'03), 2003.

[THO1992] Thompson, M. & Wishbow, N. "Improving Software and Documentation Quality Through Rapid Prototyping." In Proceedings of the 10th Annual International Conference on Systems Documentation (1992).

[SUD2005] Sudhakar M., Managing the Impact of Requirements Volatility, Master Thesis, Department of Computing Science, Umeå University, Sweden, 2005, http://www.cs.umu.se/education/examina/Rapporter/MundlamuriSudhakar.pdf.

[SAG1990] Sage, Andrew P., and Palmer, James D. Software Systems Engineering. New York: John Wiley & Sons, 1990.

3 Web Portals

In this chapter, we present the state of the art in the area of Web portals. Currently, there are two main streams in the development of Web portals

This chapter is organised as follows. First we present different approaches on web portals and compare traditional portal applications with content management systems. Secondly, we present some current standards in the development of web portals. Further, we investigate several platforms for developing web portals, including Grid-enabled portals. Finally, we investigate the state-of-the-art in the field of biomedical web portals.

3.1 Portal Technologies

3.1.1 Web Portals

Web portals are sites on the World Wide Web that typically provide personalized capabilities to their visitors. Since they are designed to deal with a great number of visitors simultaneously, the web portals use distributed applications. Different types of middleware and hardware ensure the interface between these applications and the good functioning of the portal.

The basic functionalities of a Web portal are:

- Contextualizes and frames large content sets
- Delivers personalized or customized content to audience segments or individual end users
- Manages access to published content and applications (single sign on)
- Aggregates content

The proliferation of the Web portals came together with the development of the net-browsers, in the mid 90s, when more and more companies tried to have a piece of the Internet market, by acquiring or building a Web portal. The typical services offered by Web portals are: news, interest groups, forums, chats, free email service, games etc. Important portals are: Yahoo!, AOL, Firstgov (the portal of the US government), Directgov (the UK's government portal); mini-portals are localized portals, based on local interests, that do not provide the same levels of services as major portals but they are used for collaboration of ideas, for commonly interested people (see KNET at www.silvernet.bravehost.com for example).

Most portals are written in Java, through the technology of portlets. Consequently, we will be interested in studying Java Portals and Web standards for Java portals and portlets.

3.1.2 Content management Systems

A CMS (Content Management System) is, as related to Web portals, a Web application used for managing websites and Web content (for example assisting in automating various aspects of Web publishing, like the CMSs called Wiki).

The basic functionalities of a CMS are:

- Manages the people who author content (access; rights; workflow)
- Facilitates the upholding of standards
  - A tool for applying common look and feel
- Audits the publishing environment
  - Who, what, when
- Manages the publication of content
  - Structures content for delivery

Since a large number of CMSs are available, an importat question is: “in what way one should choose an appropriate CMS for its own portal?”. This question is approached by James Robertson in the paper “How to evaluate a content management system”. The most important aspect that comes out of this paper is that a CMS should be chosen in such way that the way the steps of its life-cycle (namely Content creation, Content management, Publishing, Presentation, Contract & business) are implemented complies with the requirements of the portal. Also, the author reviews some main features that a CMS should have (for details see: http://www.steptwo.com.au/papers/kmc_evaluate/).

Most of the current CMSs are written in PHP, but there are also Java CMSs. Though, there are no standards for CMSs.

3.1.3 Web Portals vs CMSs

Web portals and CMSs have many identical features, but different teams and communities develop and maintain them. Both of them are used for building portals and Web sites. The natural question is: what are the facts that make them different (besides names)?

We can summarize the differences between a Web portal and a CMS by this definition: while CMSs are more concerned with **producing** and **managing** the content, Web portals are more concerned with **aggregating** and **delivering** the content.

However, one cannot say that CMSs are only about content production and portals only about content delivery, since both technologies have borrowed the other’s features.

3.2 Web portal standards

In order to describe the state of the art products in the area of online training platforms we must refer to the existing standards of the domain. The standards we present here mostly cover the area of Web portals.

3.2.1 Web Services for Remote Portlets

Web Services for Remote Portlets (WSRP) is a standard for Web portals to access and display portlets that are hosted on a remote server. The WSRP specification defines a web-service interface for interacting with interactive presentation-oriented web services. It has

been produced through the joint efforts of the Technical Committees of the Organization for the Advancement of Structured Information Standards: Web Services for Interactive Applications (WSIA) and Web Services for Remote Portals (WSRP). The actual version of this standard is: WSRP1.0, since August 2003. Right now, the second version of this standard WSRP2.0 is under development.

The developers of the standard motivate their work by the fact that integration of remote content and application logic into an End-User presentation has been a task requiring significant custom programming effort. It has been remarked that, typically, vendors of aggregating applications, such as a portal, write special adapters for applications and content providers to accommodate the variety of different interfaces and protocols those providers use. The goal of the WSRP specification is to enable an application designer or administrator to pick from a rich choice of compliant remote content and application providers, and integrate them with minor programming effort.

The document aims to fulfil the goal through a standard set of web service interfaces allowing integrating applications to quickly exploit new web services as they become available. The authors of the specification of the standard want to maximize the reuse of presentation-oriented, interactive web services while allowing the consuming applications to access a much richer set of standardized web services.

3.2.2 JSR 168

The second standard that we address is the Java Portlet Specification. This 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 (having as members experts from leading industry companies) as JSR 168, and released in October 2003. Right now, the second version of the standard, Java Portlet Specification V2.0 or JSR286, is developed.

The Java Portlet Specification V1.0 introduces the basic portlet programming model with:

- two phases of action processing and rendering in order to support the Model-View-Controller pattern.
- portlet modes, enabling the portal to advise the portlet what task it should perform and what content it should generate
- window states, indicating the amount of portal page space that will be assigned to the content generated by the portlet
- portlet data model, allowing the portlet to store view information in the render parameters, session related information in the portlet session and per user persistent data in the portlet preferences
- a packaging format in order to group different portlets and other J2EE artefacts needed by these portlets into one portlet application which can be deployed on the portal server.

As a reference implementation of the JSR168 standard stands the portal: http://portals.apache.org/pluto/.

JSR 168 and WSRP are standards specifications that address different aspects of portlet functionality. JSR-168 is a technology specific (Java) Portlet API designed to enable interoperability between Java portlets and Java portlet containers. WSRP is a technology agnostic protocol designed to remote Portlets in a standard manner. These two are not

orthogonal, but rather parallel. WSRP does not make any statements as to how the protocol should be implemented. These two specifications can be (and frequently are) used together with JSR 168 defining a portlet and WSRP remoting that Portlet to remote containers/consumers.

3.2.3 Other Standards

The above two standards are used along side other existing or emerging web-development standards, such as:

- WSDL – Defines how abstract interfaces and their concrete realizations are defined.
- Schema – Defines how types are defined and associated with each other.
- Namespaces – Defines how XML Namespaces are declared and used.
- SOAP – Defines how to invoke web service interfaces.
- URL – Defines URI (includes URL) syntax and encoding
- XML Digital Signatures – Defines how portions of an XML document are digitally signed.
- SAML – Defines how authentication and authorization information may be exchanged.
- XACML – Defines syntax for expressing authorization rules.
- P3P – Defines how a Producer/Portlet may publish its privacy policy so that a Consumer could enforce End-User privacy preferences.
- XML Encryption – Defines how to encrypt/decrypt portions of an XML document.
- WS-Security – Defines how document level security standards apply to SOAP messages.
- RLTC – Defines syntax for expressing authorization rules.
- WS-I.org - Defining additional profiles (e.g. Security) for use of web services standards such that interoperability is maximized.
- DIME – A lightweight, binary message format that encapsulates one or more resources in a single message construct.

3.3 Web portal development platforms

We continue by presenting two outstanding web portal development platforms, one commercial and the other one open-source.

3.3.1 Liferay (portal)

Liferay Portal framework is JSR-168 compliant and runs on almost any major application server, database, and operating system, rendering over 700 deployment combinations. It provides over 50 portlets pre-bundled and more than 20 community-contributed themes available, bringing immediate usability and accelerated development potential to portal-based internet application scenarios. Liferay was built in order to have an extensive list of features that compares with most commercial portals but without the high license fees.

As important features, Liferay provides:

- **Different Themes**, that allow the user to change the look of the portal without modifying the core code
- **Different Subthemes**, permitting the customisation of the appearance of every portlet.
- **A built-in CMS**, that offers a tool for managing the information provided in the portal.
- A built-in **connector for Central Authentication Service**, Yale's single sign on engine. Liferay can also synchronize its user list between the portal and an external data source like another database or LDAP server. A default connector for Microsoft Exchange is bundled with the portal.
- Liferay was designed to be used by application service providers. This means **hosting multiple instances** of the portal (distinguished by unique URLs) on one application server and database.
- Unlike portals that come from application server vendors, Liferay is designed to be **application server agnostic** so you are not locked into a specific server. Liferay will work on lightweight servlet containers like Jetty and Tomcat, or on J2EE compliant servers like Borland ES, JBoss+Jetty/Tomcat, JOnAS+Jetty/ Tomcat, JRun, OracleAS, Orion, Pramati, RexIP, Sun JSAS, WebLogic, and WebSphere. Being a Java portal means also that this product will work on many operating systems: BSD (FreeBSD, NetBSD, OpenBSD), Linux (Fedora, Novell), Solaris, Mac OS X, and Windows.
- Liferay uses Hibernate as the ORM tool for the persistence layer which **enables pluggable databases** (DB2, Firebird, Hypersonic, InterBase, JDataStore, MySQL, Oracle, PostgreSQL, SAP, and SQL Server).
- **Portlets, CMS content, and page layouts** can all be localized to the languages wanted. The Language portlet can be easily added to any page and the end-user can select a different localization on the fly.
- Administrators can **easily manage users, organizations, locations, and roles** through a GUI interface. Access to portlets is also restricted to users based on roles. Administrators can also specify community pages so that all users who belong to a certain group see the same page.

3.3.2 Xaraya

Xaraya is an extensible, open Source web application framework written in PHP and licensed under the GNU General Public License. As important features we enumerate: this CMS is platform independent, respects standards, scalable, flexible, modular, secure, and easy to install.

In more details, Xaraya is a web application framework that allows separation of site layout, content and logic and provides a set of tools and components for building powerful web applications. The main components of the Xaraya Framework are, as described by the developers of the product:

- A slim and well defined core API
- Powerful Block Layout templating system

- Hierarchical roles based system for user and group management
- Finely tuned and robust roles based permission system
- Plug in events and authentication
- Dynamic data providing you a choice to extend data structures with or without programatic intervention
- Flexibility and custom functionality using a range of Team Xaraya and 3rd party pluggable extensions

Xaraya runs on most platforms that support PHP, and can be used with an ever growing list of relational databases including MySQL, PostgreSQL and more recently SQLite. Separate development scenarios provide support for additional databases. Xaraya's set of tools and components allows the definition of the work environment by letting the user choose how to build and customize the web site using:

- Xaraya's tools and components, without any programming, and plugging in extra functionality with dynamic data, extensions and hooks to additional site wide functions
- Xaraya development tools, powerful APIs and reusable code to develop complex applications in a rapid development environment
- A mixture of both the above methods

These components can be used to build web applications or systems such as:

- a simple static document website
- a dynamic database driven community site
- a content management system or news publishing site
- personal blog
- a company intranet
- portals
- a specialized industry site with custom built Xaraya applications
- an enterprise level multi-site and multi-language deployment

3.4 Grid-enabled portal development platforms

3.4.1 GridSphere (portal)

The starting point of the developers of GridSphere was the desire to build a Portal, similar with the usual Web portals such as Yahoo or Amazon, that will be able deliver the benefits of Grid computing to virtual communities of researchers and scientists, providing customizable, easy-to-use, singular access points to Grids.

GridSphere, is an open-source JSR-168 compliant portal framework that is ready to run with a suite of tutorial and example portlets. One of the key elements in GridSphere is the ability for site administrators and individual users to be able to dynamically configure the content they choose to see based upon their application needs. To make content dynamic, individual components in the portal are generally engineered independently from one another. One of the key benefits to using GridSphere comes from the additional package, GridPortlets, which

provides many of the portlets needed to produce a production Grid-portal. The main benefits are, thus, the fact that the portal is JSR-168 compliant, it uses GridPortlets and Portlet Services framework. Also it has a large user-base and support community.

The technical features listed by the developers of GridSphere consists in:

- Portlet API implementation compatible with IBM's WebSphere 4.2.
- Support for the easy development and integration of "third-party portlets"
- Higher-level model for building complex portlets using visual beans and the GridSphere User Interface tag library.
- Use of Style Sheets and User Interface tags in order to allow GridSphere to be “themable”
- Flexible XML based portal presentation description can be easily modified to create customized portal layouts.
- Built-in support for Role Based Access Control separating users into guests, users, admins and super users.
- Sophisticated portlet service model that provides functionality that can be reused across multiple portlets.
- Persistence of data provided using Hibernate supports most major databases including MySQL, Postgres, DB2, HsqlDB, etc, as in the above presented case of Liferay.
- Integrated Junit/Cactus unit tests for server side testing of portlet services including the generation of test reports.
- GridSphere core portlets offer base functionality including login, logout, user and access control management.
- Full localization support in the Portlet API implementation and GridSphere core portlets supporting several languages.
- Together with the Core services (Portlet Manager Service, that provides lifecycle methods to allow portlets to be installed, removed, initialized and destroyed by authorized users, Login Service, that allows a User to be retrieved from a username and password, User Manager Service, add/remove user accounts, edit user profiles, access control service, add/remove user groups, add/remove user roles), an important role is played by the Grid Services such as:
- Credential Manager Service, used to add/remove allowed User Credentials and configure use of Credential Retrieval Service
- Job Manager Service, used for listing, starting, migrating, stopping jobs.
- Job Monitoring Service, used to specify what to monitor for any given job and archive related information.
- File Transfer Service, useful for managing and scheduling file transfers.
- Data Manager Service, used to access to data replica catalogues, and describe data with meta-data.
- Notification Service, used to Define events to be notified about, and to specify how to be notified about those events.

While the Core services are implemented via core portlets, the Grid services are implemented via Grid portlets:

- Credential Administrative Portlets. By these portlets it can be specified what credentials are permitted for use, the mappings between credential subjects and user accounts, as well as mappings to particular resources. Also, admins can view active credentials and their usage online.

- Credential User Portlets. By these portlets users may request new credential mappings to their accounts, and may retrieve and refresh credentials for later use.

- Resource Management Portlets. These are used to specify and describe Grid resources, as tools for discovering resources on the Grid, and as tools for tracking requests made to site admins for configuring or updating resources with software, etc.

3.4.2 Grace (CMS)

GRACE (Grid seArch & Categorization Engine) is a content management system designed to enable research communities to get organized around their specific common interests, and share their computational storage and knowledge resources according to their specific information needs.

The developers of GRACE state that the product is based on the principle that a content management system must not replace or unnecessarily enlarge the customer's existing resources, but rather allow the customer to maximize their use. To obtain this, two major advantages were implemented: integration of the existing content sources and Grid technology.

GRACE also implements an innovative approach to the integration of multiple content sources: it systematically harvests relevant information from these documents, applying very strong natural language processing methods in order to re-index them into Knowledge Domains. Knowledge Domain is not only a complete virtualization of multiple relevant content sources, but also incorporates the underlying semantics encapsulated in related ontologies. These ontologies are used to: query the content sources, and than index them by associating them with the key terminology; to extract from the extensive content sources only the portion that is indeed relevant for a particular interest; the ontologies are further used for querying, just like a normal index, browsing, and presentation of the search results. Moreover, it is performed a constant update of the Knowledge Domains with new and relevant information.

Also, the utilization of the strong natural language processing methods allows GRACE to offer unprecedented Information Retrieval functionalities to the knowledge workers. GRACE provides multilingual abilities for several languages, as: English, German, Italian, but other languages can be added by integrating suitable lexical databases.

GRACE utilizes the Data Grid (a special case of Grids that assume that data consumed by these computations must be shared across an organization and that it is thus more efficient to keep these data stored in a central location, accessible to various front end applications targeting various organizational information needs) in order to make the Knowledge Domains securely accessible throughout an organization, regardless of the geographical location or the strength of the locally available resources.

The developers of GRACE, state that their product is the first working system that allows integration of the unstructured, textual information.

3.5 Biomedical web portals

3.5.1 caBIG

The cancer Biomedical Informatics Grid, or caBIG, is a Grid connecting individuals and institutions to enable the sharing of data and tools, creating a World Wide Web of cancer research. The goal is to speed the delivery of innovative approaches for the prevention and treatment of cancer. The infrastructure and tools created by caBIG also have broad utility outside the cancer community. caBIG is being developed under the leadership of the National Cancer Institute's Center for Bioinformatics.

The current test bed architecture of caBIG is caGrid. The software embodiment and corresponding documentation of this architecture constitute the caGrid 0.5 release. The caGrid 0.5 software release contains tools for creating and deploying caBIG-compliant Grid services to the caGrid 0.5 infrastructure. The software infrastructure is designed to satisfy the use case requirements from the various caBIG Domain Workspaces (such as: Clinical Trial Management Systems, In Vivo Imaging, Integrative Cancer Research, Tissue Banks and Pathology Tools). caGrid 0.5 is providing the necessary infrastructure for caBIG applications to leverage the following Grid infrastructure capabilities:

- Indexing and Registry Services
- Metadata Management
- Common Data Elements
- Controlled Vocabulary Semantics
- XML Schema Management
- Security Services
- Discovery and Invocation
- Data Service Toolkit
- Analytical Service Toolkit

3.5.2 Telescience

In 1999, web-based Telemicroscopy was released and researchers were able to effectively use the remote interface to acquire data. It became clear that for a complete remote research scenario, the ability to remotely acquire data must be closely coupled to data computation and storage resources. The Telescience Project was developed to address this issue.

Telescience, provides a Grid-based architecture to combine the use of Telemicroscopy with tools for:

- Parallel distributed computation,
- Distributed data management and archival, and
- Interactive integrated visualization tools

to provide an end-to-end solution for high throughput microscopy.

The Telescience Project merges technologies for:

- Remote control,
- Grid computing, and
- Federated digital libraries of multi-scale, cell-structure data.

The objective of the Telescience Project is to increase the throughput of data acquisition and processing and ultimately improve the accuracy of the final data product.

The Telescience Portal represents the interface for the Telescience Project. It consolidates access for controlling instruments remotely, managing data, and controlling batch jobs with a single login and password. The Portal walks the user through the complex process of remote data acquisition via Telemicroscopy; Globus-enabled parallel tomographic reconstruction; advanced visualization, segmentation, and data processing tools; and transparent deposition of data products into federated libraries of cellular structure. Key features of the Portal include personalized user information, collaboration tools such as chat and shared white boards, and automatic storage of data and job tracking tools.

There are currently several portals supported under the Telescience project:

- NCMIR Multi-Scale Imaging Portal (the original Telescience-based Portal) http://swilken.ucsd.edu:8080/Gridsphere/Gridsphere
- NIEHS Environmental Health Science Data Resource Portal http://balata.ucsd.edu:8080/Gridsphere/Gridsphere
- Biomedical Informatics Research Network Portal https://portal.nbirn.net/BIRN/
- Branfman Family Foundation Collaboration Portal http://pebblebeach.ucsd.edu:8080/Gridsphere/Gridsphere
- NCI National Brain Cancer Model Collaborative Network (nBCMn) Portal http://birkdale.ucsd.edu:8080/Gridsphere

All these implementations are based on Gridsphere technology.

3.6 Initial Conclusions

We outline here only several major conclusions that can be extracted from the above presentation.

- Web portals and especially the ones using opensource technologies are based on standard Java portlets.
- Sometimes web portals are replaced by Content Management Systems (CMSs).
- Web portals and CMSs have many common features, but, basically CMS are more concerned with the production of content, while Web portals are more concerned with the delivery of content.

3.7 References

[WEBOPEa] Opensource Content Management Systems basically written in PHP and MySQL HTTP://WWW.OPENSOURCECMS.COM/

[WEBJAVa] Opensource Content Management Systems written in Java HTTP://JAVA-SOURCE.NET/OPEN-SOURCE/CONTENT-MANAGMENT-SYSTEMS

[WEBJAVb] Opensource Web Portals written in Java HTTP://JAVA-SOURCE.NET/OPEN-SOURCE/PORTALS

[WEBSTE] J. Robertson, How to evaluate a content management system HTTP://WWW.STEPTWO.COM.AU/PAPERS/KMC_EVALUATE/

[WEBWEL] L. Welchman, CMS vs Portal – Which one do you need, HTTP://WWW.WELCHMANCONSULTING.COM/PRESENTATIONS/CMS_PORTAL.PPT

[WEBOPEb] Opensource Grid-enabled Web Portals HTTP://WWW.OPENGRIDPORTALS.ORG/

[WEBGRI] Gridsphere Webpage HTTP://WWW.GRIDSPHERE.ORG/

[WEBGRA] Grace Webpage HTTP://WWW.GRACE-IST.ORG/

[WEBLIF] Liferay Webpage HTTP://WWW.LIFERAY.COM/

[WEBXAR] Xaraya Webpage HTTP://WWW.XARAYA.COM/

[WEBCAB] The caBIG Portal HTTPS://CABIG.NCI.NIH.GOV/

[WEBTEL] The Telescience Project HTTP://TELESCIENCE.UCSD.EDU/

PART 2

Portal User Scenarios

4 ACGT Portal users

ACGT Portal users are all the users of the ACGT infrastructure. Following the document D2.1 “User requirements and specification of the ACGT internal clinical trial”, the types of ACGT users can be described upon several criteria.

We will investigate in this chapter the relevance of ACGT stakeholders and users against the ACGT Portal usability.

4.1 ACGT stakeholders

The ACGT Project will ultimately impact anyone who is involved in post-genomic clinical research and clinical trials on Cancer. A closer analysis reveals that it appears to encompass a significant number of stakeholders. Stakeholders are defined as:

> *“An enterprise, organization, or individual having an interest or a stake in the outcome of the engineering of a system.*”

This should not be a surprise given the importance of the domain of application, i.e. post-genomic cancer research form a large number of organisations. Indeed, the breadth of the project spans the universe of post-genomic, individualised medicine and information technologies, while the classes of stakeholders are quite large and diverse with respect to areas of interest, depth and domains of expertise, and desire for involvement.

Key representative users are members of the ACGT Consortium, whereas others will be drawn into the ACGT user community as the project progreses.

The key stakeholder groups we see in ACGT are briefly described in the following sub-sections.

4.1.1 Cancer Research Organisations

The Cancer research community depends on data located in multiple disparate data sources around the world. In most cases, these data sources do not utilize the same underlying data structure. However, given these tremendous barriers to their research and discovery efforts, the cancer research community has been able to perform a wide range of studies to elucidate the mechanisms and treatment of cancer.

Most participants in the cancer community have access to the internet and use protocols such as file transfer protocol (FTP) and hypertext transfer protocol (HTTP) to access disparate data sources. As the community grows, many participants have become blind to the seemingly endless number of new datasets, methods and tools published on a daily basis.

Cancer research organization, members or not of the ACGT Consortium, should play an active role on the ACGT Portal.

Besides the ACGT Consortium members that fully agreed to share their biomedical and clinical resources under the legal and ethical safety provided by the ACGT infrastructure, the

project aims to attract in the ACGT system also other organizations that are not willing, at least from the beginning to fully integrate their resources in the ACGT system.

The ACGT Portal can play an essential role by integrating resources that are not distributed under the ACGT Grid System, but provide value in the cancer research and enrich the world-wide collaboration.

Thus, the ACGT Portal will incorporate both ACGT grid enabled resources, which will benefit from all the facilities offered by the grid architecture (single sign-on, speed, availability, etc) and also "external" resources that will be marked correspondingly and will be available according the policies maintened by each owner.

4.1.2 Researchers and Scientists involved in post-genomic research

The lack of a unifying architecture has proven to be a major roadblock to a researcher's ability to mine different databases. Most critically, however, even within a single laboratory, researchers have difficulty integrating data from different technologies because of a lack of common standards and other technological and medico-legal and ethical issues. As a result, very few cross-site studies and clinical trials are performed and in most cases it isn't possible to seamlessly integrate multi-level data (from the molecular to the organ, individual and population levels).

Moreover, clinicians or molecular biologists often find it hard to exploit each other's expertise due to the absence of a cooperative environment which enables the sharing of data, resources or tools for comparing results and experiments, and a uniform platform supporting the seamless integration and analysis of disease-related data at all levels and simulating the mechanisms of disease evolution.

Researchers and scientists are the main group of users of the ACGT Portal that are not only data consumer, but also data producer, since their activity in the portal will produce new workflows, results of the investigations, relevant literature and the emulation that the portal needs to live.

4.1.3 Technology Suppliers

Technology stakeholders are interested in designing, building, integrating, and servicing products that would effectively become a part of the implementation of the IECSA.

These individuals would be adopters of the architecture specifications and associated standards. As such, vendors provide a different view of the IECSA requirements – one clearly focused on the implementation and long term use of the architecture. The purpose of engagement with vendors is to gain acceptance of the project concepts from vendors and suppliers directly designing, building, and supplying equipment. Many vendors already participate in standards bodies and this project must be presented in the context of building upon existing standards development work.

Such manufacturers are likely to have substantial technical input that must be sought as the technical requirements are being established.

However, it is unlikely that all the technology suppliers will be able to meet the standards immediately. In this respect, the ACGT Portal can link resources that cannot be fully integrated under the ACGT Grid System and Ontology, but that are required by the research community.

Moreover, the ACGT Portal can implement separately for each supplier the access policy to the supplier's resources.

4.1.4 Patients and Patient Organisations

In greater Europe, there are 2.8 million new cases of cancer every year, and 1.7 million die from the disease every year. In the EU 25 alone, 4.5 European citizens have had or have cancer. There are 2 million new cases every year, and 1.16 million die from it every year (2004).

The growing cancer burden in Europe can be illustrated in many other ways, each telling of individual suffering and loss, but also of a need to act collectively to eliminate this scourge:

- 1 in 3 men and 1 in 4 women will be directly affected by cancer in the course of the first 75 years of their life.
- 3000 people die from cancer every day in Europe.
- Cancer is the **second main cause of death** in Europe, after circulatory diseases.

By complementing public actions towards better health and by collaborating with local and European authorities to fight cancer it is of utmost importance to coordinate this lobbying activities in close cooperation with patients and patient cancer organisations. Today for nearly every cancer a patient organisation is known. In the European context patients' coalitions did form Cancer leagues, to provide support to cancer patients and their relatives, and to improve the quality of treatments.

The **Association of European Cancer Leagues** (ECL) is a federation of national and regional Cancer Leagues (http://www.europeancancerleagues.org/), made of either patients' coalitions or of cancer control professionals. The objectives of the association are to improve communication, to promote, enhance and co-ordinate collaboration between European leagues/societies and to foster fruitful activities between European cancer leagues and organisations, in order to reduce the growing cancer burden in Europe.

ECL is located in Brussels and is a non-for-profit association (asbl; association sans but lucratif), under Belgian law. ECL was created in 1980, and consists of 31 members today, located all over extended Europe (See Appendix 5 for a listing and details of members).

Regarding Paediatric Cancer the **International Confederation of Childhood Cancer Parent Organisations** (ICCCPO) was founded in May 1994 in Valencia, Spain. ICCCPO is a worldwide network of organisations of parents of children with cancer. The mission of ICCPO is to share information and experience in order to improve access to the best possible care for children with cancer everywhere in the world.

ECL as well as ICCCPO will be contacted by ACGT asking for collaboration and for providing names of persons, who will attend Management Board Meetings of ACGT. These representatives of the patient organisations should especially be enrolled in the discussion of legal and ethical issues. The goal is to strengthen the collaboration between patients and ACGT to foster fruitful activities in order to bring basic research faster into clinical practice.

The role of patient organizations is to monitor the activities that take place through the ACGT Portal, to be aware of all new trials that are performed and to speed up the adoption of the results provided by these trials in the clinics. Patients and patient organizations may use the ACGT Portal in order to communicate with the scientific community.

4.1.5 Regulatory Agencies

Regulators and auditors have an interest in ensuring that the ACGT systems meet their reliability, performance, market, and financial obligations.

The purpose of this engagement is to assist regulatory commissions in understanding the nature and need for a project that develops an industry-wide architecture and the problems that arise from lack thereof.

4.1.6 Standards Bodies

The purpose of engagement with these groups is to gain acceptance and future standardization of the concepts. Engagement with standards groups such as the IEEE, GGF, and others is planned.

Also, clinical trials related standards bodies, such as the International Conference on Harmonisation (ICH) which is a clinical trials related standard. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) is a unique project that brings together the regulatory authorities of Europe, Japan and the United States and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration.

The purpose is to make recommendations on ways to achieve greater harmonisation in the interpretation and application of technical guidelines and requirements for product registration in order to reduce or obviate the need to duplicate the testing carried out during the research and development of new medicines.

The objective of such harmonisation is a more economical use of human, animal and material resources, and the elimination of unnecessary delay in the global development and availability of new medicines whilst maintaining safeguards on quality, safety and efficacy, and regulatory obligations to protect public health.

Some issues that standards bodies and industry consortia can expand upon for the ACGT team include:

- Drivers to harmonize standards in progress
- Needs to address standards integration on enterprise and industry levels
- Integration of standards initiatives and past work
- Drivers and needs to establish interworkability testing and other formal methods of integration
- Integration of architecture methods with standards development activities

Standard bodies can assist the implementation of new standards within the ACGT infrastructure and use the ACGT Portal

4.1.7 Market Participants

Market participants, as part of the User Community, have unique communication requirements, reflecting their market-driven needs for timeliness, availability, and security of

many different types of information. These requirements are not always met, particularly as the market environment, policies, and capabilities change over time.

Market participants can be involved in the community of the ACGT Portal by the exchange of information from the ACGT community to the market (news, literature, etc) and from the market to the ACGT community (polls, feedback forms, etc).

4.2 ACGT end-users

Regarding end-users of the ACGT infrastructure one can further distinguish:

- *Clinicians*
- *Biomedical researchers*
- *Data miners*
- *Patients*

4.2.1 Clinicians

Clinicians typically want to apply very standardized procedures. For instance, given a microarray result for a patient, they want to apply well known predictors combining clinical and microarray data (e.g. St-Gallen or NIH consensus in breast cancer) and get the results in the clearest possible fashion. Simple interfaces (e.g. web-based) should usually be sufficient for this category of users.

4.2.2 Biomedical researchers

The category of biomedical researchers includes typically PhD students, post-docs and senior scientists working in a clinical or research environment. Such users typically want to have a greater flexibility in the visualization of their data, and to have the opportunity to use various statistical tests and tools, but usually on a single set of clinical data. Interactivity in the tools is essential for these users (e.g. ability to click on a gene to get information about it, to filter their data based on selecting one branch of a cluster, to visualize gene expression as a function of the position on the genome [which typically requires zooming capabilities], to build Venn diagrams of gene lists found significant under different conditions, etc…). To some extend, these users will develop automatic methods (workflows) to complete their jobs. These WFs will be used by inexperienced users

4.2.3 Data Miners

The category of data miners overlaps somewhat with the previous category (e.g. some biomedical researchers may be willing to combine their findings with those of other studies). Data miners are typically biostatisticians willing to extract knowledge from a study using statistical tools they just developed or from a combination of studies or both. Meta-analysis tools would be of interest for that category of users. The tools used are mainly programs and scripts (e.g. in R), with less strong need for interactive graphics.

4.2.4 Patients

The role of the patient would be limited to access his/her clinical record, and to obtain various summaries of the findings associated to his/her clinical data. From the viewpoint of the ACGT infrastructure patients may be considered as clinicians with an access limited to a single subset of data (their own). As the ACGT infrastructure should essentially be a platform for improving clinical data knowledge mining, we recommend that the focus of the development effort to be put on the first three categories of users. Patients access to ACGT should be kept for a later phase of development of the infrastructure.

4.3 ACGT access levels

From the viewpoint of access to the components of the ACGT infrastructure, we identify the following actors:

- System administrators
- Service developers
- Content reviewers
- Registered users
    - Data producers
    - Data consumers

4.3.1 System administrators

System administrators are selected users with permission to manage the internals of the system. Administrators will be nominated by the ACGT Consortium and will be responsible for the good functioning of the system.

ACGT Portal administrators will probably be different from the ACGT Grid Platform administrator having specific tasks in the management of the portal. Portal administrators have technical background and they are less concerned with the content hosted by the portal.

4.3.2 Service developers

Service developers are the users which provide new applications. A controlled and supervised registry of such new services should be provided, including a version control mechanism.

It is unlikely that service developers will have direct access for adding or upgrading their applications in the ACGT Portal, but they will be able to add news regarding their applications and to visualize statistics and gather feedback about the use or their applications through the ACGT Portal.

4.3.3 Content reviewers

Content reviewers represent a special category of users specific to the ACGT Portal. Content reviewers have limited rights of administration in the portal, but they are related to the administration of content rather than to technical issues.

The administration of a portal is more efficient if it is split on areas of interest and each area is managed by a directly interested person. These area of interest can start from a homepage and end to a section in the portal.

Content reviewers have the role of:

- approving the content,
- adding new content,
- updating the content,
- moving old content in archives,

all these actions, of course, in their area of interest.

4.3.4 Registered users

As a general rule, all the users of the ACGT Portal will have to register and authenticate before being able to access the content provided by the portal.

From the data point of view, registered users can be split in data producers and data consumers. The same user can be simultaneously data producer and consumer, but normally one of these characteristics will be dominant.

4.3.4.1 Data producers

Data producer is any user that adds some resource in the portal regardless the type of this resource.

This could be a clinician who “owns” the data collected in the context of his/her clinical study. The data owner should be the only one authorized to modify his data. It is needed to identify the ownership of data submitted to the ACGT system.

A data producer can be a researcher, which develops trials through the portal and produceses new methods (workflows) of investigation. Results of the trial, scientific papers and reports are other data generated by researchers in the portal.

To some degree, any user that posts news or messages in the forum can be seen as a data producer.

4.3.4.2 Data consumers

Data consumer is any user that exploits the content of the portal in some way.

This could be any clinician who does not participate in a trial but is interested on getting some data from it. This role involves implementing a policy of access to trial data. Data

storage, data distribution and microarray data have to be in accordance to ethical and legal issues.

Other typical data consumer are:

- The patients, which browse the portal for news and other general information.
- The representatives of cancer research organizations, of regulatory organizations, or of standard bodies.

5 ACGT Portal User Scenarios

In this chapter we will describe the sequence of steps that each type of users is allowed to perform in the ACGT Portal.

The privilegies shared by the users of the ACGT Portal can be summarized by the following operations:

- Read an item
- Run an item
- Submit an item
- Approve an item
- Modify an item
- Archivate an item
- Delete an item

Fot safety reasons (legal, ethical, etc.) we will restrict the delete operation only to the administrators of the ACGT Portal. The first 5 operations will characterize the full-access that a user may have in a section of the portal.

Moreover, the aproval operation is restricted to special groups of users: editors, administrators. These groups must ensure not only that the information presented in the portal respects the scientific standards defined, but also that the legal, ethical and security aspects involving the items that are presented/discussed in the portal.

The user scenarios presented in this chapter describe the roles that are active in the ACGT Portal. An individual user can play simultaneously several roles, like reviewer and biomedical researcher.

5.1 Role: Administrator

The **Administrator** is a role that has full rights in all respects. The administrator has all privilegies, including the role of defining new roles, groups of users, attaching users to groups or defininig new sections in the portal. Also, the Administrator should deal with security issues regarding the system (such as protection of the users identity, protection of the data published, etc).

The role of administrator is pre-defined and built-in the system.

5.2 Role: Editor

The **Editor** is a role that cannot be selected as one of the default profiles, although the log-in and interaction between this user and the portal is not modeled separately.

The main role of an editor is to keep is to keep the content in his editorial area in an accurate state, regarding completeness, correctness, relevance and up-to-the-date. Also, the editor should deal with ethical, legal and security issues regardin the content, as it was suggested above.

The editor should be able to access the queue of information, data, research papers, tutorials and courses submitted to be published on the portal in order to approve or reject these items in/from the portal. The editor decides where an accepted item can be published in the portal and how long this item shall be active in the portal or should be send to the archives of the portal.

A regular user could become editor when the administrators of the project select him/her.

The editor's interaction in the portal should be modeled as follows:

1. The editor logs-in.
2. The default privilegies for these users should include:
    - Full access to the queue of the items submitted in the content area where the editor is encharged.
    - Full access to the database containing clinic data and information.
    - Full access to the database of research papers.
    - Full access access to the workflows that are related to the content area where the reviewer is encharged.
    - Full access to the forum, where the editor should be able to monitor all the discussions about the data and information presented in his editorial area.

    Each section should be accessed by a link from the main page of the portal (after log-in)
3. The editor selects an item from his editorial area.
4. The editor distributes a new item to a reviewer to make the accept/reject decision, collects the reviews from the reviewers, accepts or rejects new items (based on reviews), establishes the sections where a new accepted item can be published in the portal or sends an old item to the archive.
5. During the editing procedure, the editor moderates the discussions in the forum corresponding to his editorial area.
6. During the editing procedure, the editor should be able to access all the other resources hosted by the portal.

7. The editor logs-out.

5.3 Role: Reviewer

In order to respect the scientific and ethical standards imposed in the portal, a special type of user should be implemented.

The **Reviewer** is a user that cannot be selected as one of the default profiles, although the log-in and interaction between this user and the portal is not modeled separately. The reviewer should be able to access the queue of information, data, research papers, tutorials and courses submitted to be published on the portal. Each such record should be reviewed by at least two/three reviewers and added to the portal in the case it is considered interesting to the other users, and respects the standards regarding the structure of the information published. The reviewer should provide feedback to the submitters in a regular journal-like fashion.

Moreover, a reviewer should follow and be informed at any moment on the current state of the standards proposed on the portal, on the emerging standards, and the way they will be used. Another job of the reviewer is to check if whenever a new standard is released, the data and informations submitted before that moment is updated.

A regular user could become reviewer when the administrators of the project select him/her.

Since in many cases reviewing a submitted item implies browsing the already existing records, the reviewer should have full access to the database, to the database of research articles, and to the discussion forum regarding these articles.

The reviewer's interaction in the portal should be modeled as follows:

1. The reviewer logs-in.
2. The default privilegies for these users should include:
    - Read/Modify access to the queue of the items submitted in the content area where the reviewer is encharged.
    - Read access to the database containing clinic data and information.
    - Read access to the database of research papers.
    - Run access to the workflows that are related to the content area where the reviewer is encharged.
    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the data and information presented in the site.

    Each section should be accessed by a link from the main page of the portal (after log-in).
3. The reviewer selects an item from the submitted items queue.

4. The reviewer reviews this item, provides information on his/her review, and decides if this item is accepted for publication in this form, needs to be revised (due to scientific reasons, does not respect the standard form, or any other reason), or is rejected. The final decision is communicated to the user that submitted the item, together with the review by the editor of the area in which the item was submitted.

5. During the review, the reviewer should be able to access all the other resources hosted by the portal.

6. The reviewer logs-out. In the case that a review is in progress, the portal should save the work done so far, and permit the reviewer to continue his/her work.

5.4 Role: Cancer research organizations representative

The research cancer organizations **Representative**s are interested mainly in seeing that the data provided by their organization is well treated in the ACGT Portal. They would want also to measure the impact and relevance that this data has in the research community.

In this spirit, the portal will offer to these users the possibility of getting informed on the activity that is done on the portal regarding their own resources.

The portal will provide these users with:

- Statistics regarding the database (number of queries, number of added records, number of distinct users that accessed the database, number of distinct users that contributed to the database, the statistic of each type of users that accessed the database, etc.).

- Statistics regarding the research papers.

- Statistics regarding courses and tutorials.

The representative's interaction in the portal should be modeled as follows:

1. The user logs-in.

2. The default facilities for these users should include:

    - Read access to statistics regarding the data-base.

    - Read access to statistics regarding courses and tutorials.

    - Read access to statistics regarding the research papers.

    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the data and information presented in the site.

3. The user reads and prints the statistics.

4. The user logs-out.

5.5 Role: Researcher

We assume that **Researcher**s involved in post-genomic research, are well trained users, familiar with the state-of-the-art research in the domain.

The main problem in cancer research, as identified in the user requirements chapter, is the access to data and information provided by other organizations in an integrated manner. Also, it is stated that the main difficulty that researchers have is integrating data from different technologies because of a lack of common standards and other technological and medico-legal and ethical issues.

These specialists would want to perform the following actions:

- Access, review, comment, search, download the data contained in the underlying "database" (this data-base contains clinic data and informations) of the portal.
- Access, read, comment, search, download state-of-the-art research papers in the domain.
- Access, read, comment, search, download courses and tutorials provided by other organizations
- Read and be informed on the last data and information or research papers that were published on the portal; also, the user should be informed on the latest important news in the domain (the portal should keep a list of the news not read by the user)
- Be informed on the standards respected by the data and the information provided on the portal. A document regarding these standards should be provided to the users that access the portal and are involved in research.

The researcher's interaction in the portal should be modeled as follows:

1. The user logs-in.
2. The default facilities for these users should include:
    - Read access to the data-base containing clinic data and information.
    - Read access to the data-base of research papers.
    - Read access to the tutorials and courses offered.
    - Run access to the workflows.
    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the data and information presented in the site.
    - Read access the news, and the list of the last published records.
    - Read access to the document containing the standards that are respected on the portal.

    Each section should be accessed by a link from the main page of the portal (after log-in).

3. The user reviews the standards of the portal.

4. The user performs queries, accesses, downloads, discusses (in the forum) the records in the data-base with clinical information. The user updates, if possible, the data submitted before, that do not respect the last version of the standards.

5. The user searches the research papers, reads, downloads such papers. The user should be able to ask questions on, discuss, and cite research papers on the forum.

6. The user accesses any online tutorials or courses, download the course material.

7. The user accesses, reads, cites the latest news; accesses the latest records inserted in the databases.

8. The user submits research papers, clinical data and information, courseware.

9. The user logs-out.

5.6 Role: Clinician

Clinician is a role that is very similar to researcher. The main difference between the clinician and the researcher is that the clinician is less interesting in developing new methods and techniques and is rather interested in comparing his own data with data already existent in databases and to establish correct diagnosis and genomic profile of their patients.

The clinicians would want to perform the following actions:

- Access, search, download the data contained in the underlying "database" (this data-base contains clinic data and informations) of the portal.

- Access, read, comment, search, download state-of-the-art research papers in the domain.

- Access, read, comment, search, download courses and tutorials provided by other organizations.

- Contribute to the database with clinical evidences (that do not offense any ethical aspects, and respect the standards of the portal) and add research papers (free, not restrained by copyright rights). Whenever a new version of the standards package is released the user is asked to modify the data that was submitted so far by that user is modified, if possible, in such manner that it respects the new standards.

- Contribute with tutorials, courses regarding research and clinical procedures in the domain.

- Read and be informed on the last data and information or research papers that were published on the portal; also, the user should be informed on the latest important news in the domain (the portal should keep a list of the news not read by the user).

- Be informed on the standards respected by the data and the information provided on the portal. A document regarding these standards should be provided to the users that access the portal and are involved in research.

The clinician's interaction in the portal should be modeled as follows:

1. The user logs-in.
2. The default facilities for these users should include:
    - Full access to the data-base containing clinic data and information.
    - Full access to the data-base of research papers.
    - Full access to the tutorials and courses offered.
    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the data and information presented in the site.
    - Run access to the workflows.
    - Read access to the submission system: each researcher-user should be able to submit data, information research papers, tutorials, courses.
    - Read access the news, and the list of the last published records.
    - Read access to the document containing the standards that are respected on the portal.

    Each section should be accessed by a link from the main page of the portal (after log-in).
3. The user performs queries, accesses, downloads, discusses (in the forum) the records in the data-base with clinical information. The user runs workflows by providing necessary input data.
4. The user searches the research papers, reads, downloads such papers. The user should be able to ask questions on, discuss, and cite research papers on the forum.
5. The user accesses any online tutorials or courses, download the course material.
6. The user accesses, reads, cites the latest news; accesses the latest records inserted in the databases.
7. The user logs-out.

5.7 Role: Technology supplier

The technology **Supplier**s are interested in designing, building, integrating, and servicing products that can effectively serve in implementing strategies for fighting against cancer.

In our point of view, these specialists would want to perform the following actions:

- Access, read, comment, search, download state-of-the-art research papers in the domain; get informed in the latest discoveries in the domain in order to be able to build technology that implements these discoveries in real life.

- Access, read, comment, search, download courses and tutorials provided by other organizations; we assume that in order to build technology that solves cancer-related problems, the technology supplier should be able to understand the problems that the device it builds will be faced to.

- Contribute with tutorials, courses regarding the technology provided.

- Read and be informed on the last data and information or research papers that were published on the portal; also, the user should be informed on the latest important news in the domain (the portal should keep a list of the news not read by the user).

- Be informed on the standards respected by the data and the information provided on the portal. A document regarding these standards should be provided to the technology suppliers that access the portal, in order to build devices that will output data respecting these standards.

We assume that technology suppliers are not necessarily interested in the clinical data and information, hence we do not offer them default access to the clinical information data-base.

The supplier's interaction in the portal should be modeled as follows:

1. The user logs-in.

2. The default offer for these users should include:

    - Read/Submit access to the database of research papers.

    - Read/Submit access to the tutorials and courses offered.

    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the data and information presented in the site.

    - Read/Submit access the news.

    - Read access to the document containing the standards that are respected on the portal.

    Each section should be accessed by a link from the main page of the portal (after log-in).

3. The user searches the research papers, reads, downloads such papers. The user should be able to ask questions on, discuss, and cite research papers on the forum.

4. The user accesses any online tutorials or courses, download the course material.

5. The user accesses, read, cite the latest news; access the latest records inserted in the data-bases.

6. The user submits news or papers regarding the produced devices, courseware.

7. The user logs-out.

5.8 Role: Auditor

Auditor is a representative of a regulatory agency.

Such agencies are not interested in the information provided on the portal. They are interested mainly in seeing that the activity meets the performance expected and self-assumed by the authors.

In this spirit, the portal will offer to these users the possibility of getting informed on the activity that is done on the portal. The portal will provide these users:

- Statistics regarding the database (number of queries, number of added records, number of distinct users that accessed the data-base, number of distinct users that contributed to the database, the statistic of each type of users that accessed/contributed to the database, etc.).
- Statistics regarding the research papers.
- Statistics regarding courses and tutorials.
- Other statistics (number of log-ins, statistics regarding the forum etc.).

The auditor's interaction in the portal should be modeled as follows:

1. The user logs-in.
2. The default facilities for these users should include:
    - Read access to statistics regarding the database.
    - Read access to statistics regarding courses and tutorials.
    - Readaccess to statistics regarding the research papers.
    - Read access to other statistics.
    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the data and information presented in the site.
3. The user reads the statistics.
4. The user logs-out.

5.9 Role: Standard bodies representative

As stated in the user requirements chapter, the involvement of standard bodies **Representative**s in the project is important in order to gain acceptance and future standardization of the concepts. Hence, such users are mostly interested in the standards defined, in the discussions linked to these standards, in studying emerging standards (which we assume that are presented in research papers.). Moreover, we assume that the user that

are involved in developing new standards will also propose procedures and training material on the standards they proposed.

The portal will offer as the following facilities to these users:

- Access, read, comment, search, download state-of-the-art research papers in the domain.
- Access, read, comment, search, download courses and tutorials on standards provided by other users.
- Contribute with tutorials, courses regarding standardization in the domain.
- Read and be informed on the last data and information or research papers that were published on the portal; also, the user should be informed on the latest important news in the domain (the portal should keep a list of the news not read by the user). We assume that such information is important for deciding if a proposed standard is acceptable or not.

The representative's interaction in the portal should be modeled as follows:

1. The user logs-in.
2. The default facilties for these users should include:
    - Read access to the data-base of research papers.
    - Read access to the tutorials and courses offered
    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the standards presented in the site
    - Read/Submit access the news, and the list of the last published records
    - Read/Submit access to the document containing the standards that are respected on the portal.

    Each section should be accessed by a link from the main page of the portal (after log-in).
3. The user reviews the standards of the portal.
4. The user makes observations on the standards proposed so far on the portal. The user proposes enhancements of these standards.
5. The user proposes courses and tutorials on standards and standardization issues.
6. The user accesses, reads, cites the latest news.
7. The user logs-out.

5.10 Role: Patient

Patient and patient organizations representatives have mainly a reading role in the ACGT Portal.

These users would want to be informed on legal and ethical aspects of the research and to survey the way in which the rights of the patients are respected.

Also, these users would want to boost the application of the new treatment techniques from research laboratories to clinics, as long as these methods have passed all the checking and approval steps.

The patient's interaction in the portal should be modeled as follows:

8. The user logs-in.
9. The default facilties for these users should include:
    - Read access to the database of research papers.
    - Read/Submit access to the forum, where they should be able to discuss, in an organized manner, the standards presented in the site.
    - Read access the news.
    - Read access to the document containing the standards that are respected on the portal.

    Each section should be accessed by a link from the main page of the portal (after log-in).
10. The user accesses, reads, cites the latest news, papers or standards.
11. The user may ask specialists about different legal, ethical or scientifical questions.
12. The user logs-out.

PART 3

Functional specifications

6 Main functional requirements

In this chapter we try to establish which of the general features of a Web portal are of a particular interest for the ACGT Portal.

We consider the features of the ACGT Portal with regards to the two main goals of a portal:

- To deliver content
- To manage content

Speaking about **content delivery**, we have identified the following ACGT relevant features:

- User personalization

  The ACGT Portal is expected to link a huge quantity of information and resources of different types. The information linked by the portal should be delivered to users in the most accessible way.

  The first key feature meant to help users accessing exactly the information and resources they want is the personalization of the presentation layer of the portal, i.e. the way in which the information contained in the portal is presented to each user.

- Search methods

  The second key feature meant to help users to extract information from the ACGT Portal is represented by the search methods. Search methods should provide a complete and refined scan of the portal. Search tools (as well as data mining or other advanced tools) are not going to be developed in the portal itself, they will be developed as planed within other WPs, but they should be integrated in the ACGT Portal, at the interface level.

  The simple organization (as smart as this organization would be) of the information contained in the ACGT Portal is not relevant for all types of users and is going to be useless for most of them. The main method of getting information from the portal should not be browsing, but searching.

- User activity repository

  Much of the users activity in the portal represents repetition. An important facility of the ACGT Portal would be to allow the users to find their past activity in the portal (for a maximal number of days or operations) and to provide them with the support for recovering the results of their past tasks launched through the portal or for repeating these tasks, perhaps with modified input data.

  For example, the user can launch a complex task (like a workflow), then disconnect and come back after 6 hours or 2 days and find the results of the workflow run.

  The same facility might be extended to groups of users working on a specific clinical trial and become an important support feature for collaboration (see below other issues on users collaboration).

- User online assistance

  The fourth key feature of the ACGT Portal is related to the online training of the users (activity that is also comprised in WP14). The training that will be provided through the ACGT Portal should be integrated to such a degree with the activity of the user in the portal that the user will be able to learn “HOW TO” get something or “HOW TO” perform a task not by switching to a training platform but exactly in the place and at the time they need this information.

Regarding **content management**, we have identified the following ACGT relevant features:

- Detailed scheme of privilegies and roles

  The fifth key feature of the ACGT Portal concerns the development of a strong scheme of privilegies and roles. Who has the right to read, edit or modify what and when ? The first step for establishing a good content management system is to define who is responsible for each slice of the portal. The roles scheme has to be permissive enough to encourage users to share content and resources through the ACGT Portal.

- Data and resource security

  Sensitive data, like pseudonimized data on clinical cases will be manipulated through the ACGT Portal. On the other hand, data or resources produced in clinical trials that is not yet tested and confirmed will be manipulated through the portal within restricted groups of clinicians and researchers. The sixth key feature of the ACGT Portal is data and resource security, meaning that the access to the data and resources should be 100% modelled upon the scheme of privilegies and roles.

- Collaborative work

  Once users have rights to contribute to the portal, the seventh key feature of the ACGT Portal is to ecncourage the collaboration through the portal. Collaborative tools should allow users to share files, actions and messages.

In the sequel, we shall discuss in detail each of these key features of the ACGT Portal.

6.1 User customization

The first important feature for the ACGT Portal concerns the way in which the information is presented to each user.

6.1.1 Log-in

As a rule, all the content provided by the ACGT Portal, excepting for general news, will be available only for registered users.

This restriction is intended to:

- Present the content in a customized form for each user
- Provide users with all the facilities offered by the ACGT Portal

- Track users activity in the portal in order to produce statistics on which resources are more accessed, through which paths resources are accessed or how helpful is the portal for users' tasks.

Another important characteristics of the ACGT log-in functionality shall be the module of Single Sign-On. This module would allow the users to navigate through all the applications and databases integrated in the portal without needing separte authentication for each of these resources, which are available on the WEB.

However, only resources fully integrated in the ACGT Grid Infrastructure will benefit of the Single Sign-On facility. Other resources linked in the ACGT Portal, but which are not integrated in the ACGT Infrastructure will probably need a separate authentication step (probably a separate account name and password).

On the other hand, it is possible that some suppliers would require an additional password (or PIN) for the services they provide in the ACGT Portal or at least for some special (advanced, customized) parte of their services.

A visual prototype of the log-in page is presented in **Figure 4:** ACGT Portal – Login page prototype.

6.1.2 Content presentation

The ACGT Portal should present the content in a stomized manner according to the user profile and preferences. This is first of all a security feature, since any user should be able to access only the content that is allowed to. These differences, that in some cases are huge in terms of privilegies (as presented in Chapter 5), have also an important influence in the form in which the content is presented for each type of users.

The connection between the user and the privilegies that he/she has in the portal is first of all made by the roles defined in Chapter 5. According to the matrix of privilegies, the user receives at the registration a standard configuration of his/her profile determining the way in which the content is presented.

On the other hand, the users, even playing the same role in the ACGT Portal may have different preferencies regarding the presentation of the content to which they have access. Some part of this content may have no relevance for a specific user, thus he/she might want to hide permanently this kind of content. Other part of the content can be very important for the user and in this case, he/she would want to receive this content in the main window of his/her homepage in the portal.

A visual prototype of the My profile page is presented in **Figure 5:** ACGT Portal – User's homepage prototype.

The homepage of the user represents the first page in which the user is redirected after authentication, but in big lines represents the way in which the portal content which is accessible to the user is also presented to him/her.

The content is grouped in categories. Each category is accessible from the horizontal menu. However, some categories are presented directly in the current page. The content of these pre-presented categories is published in separate blocks.

Each block can be displayed in two shapes: large and small.

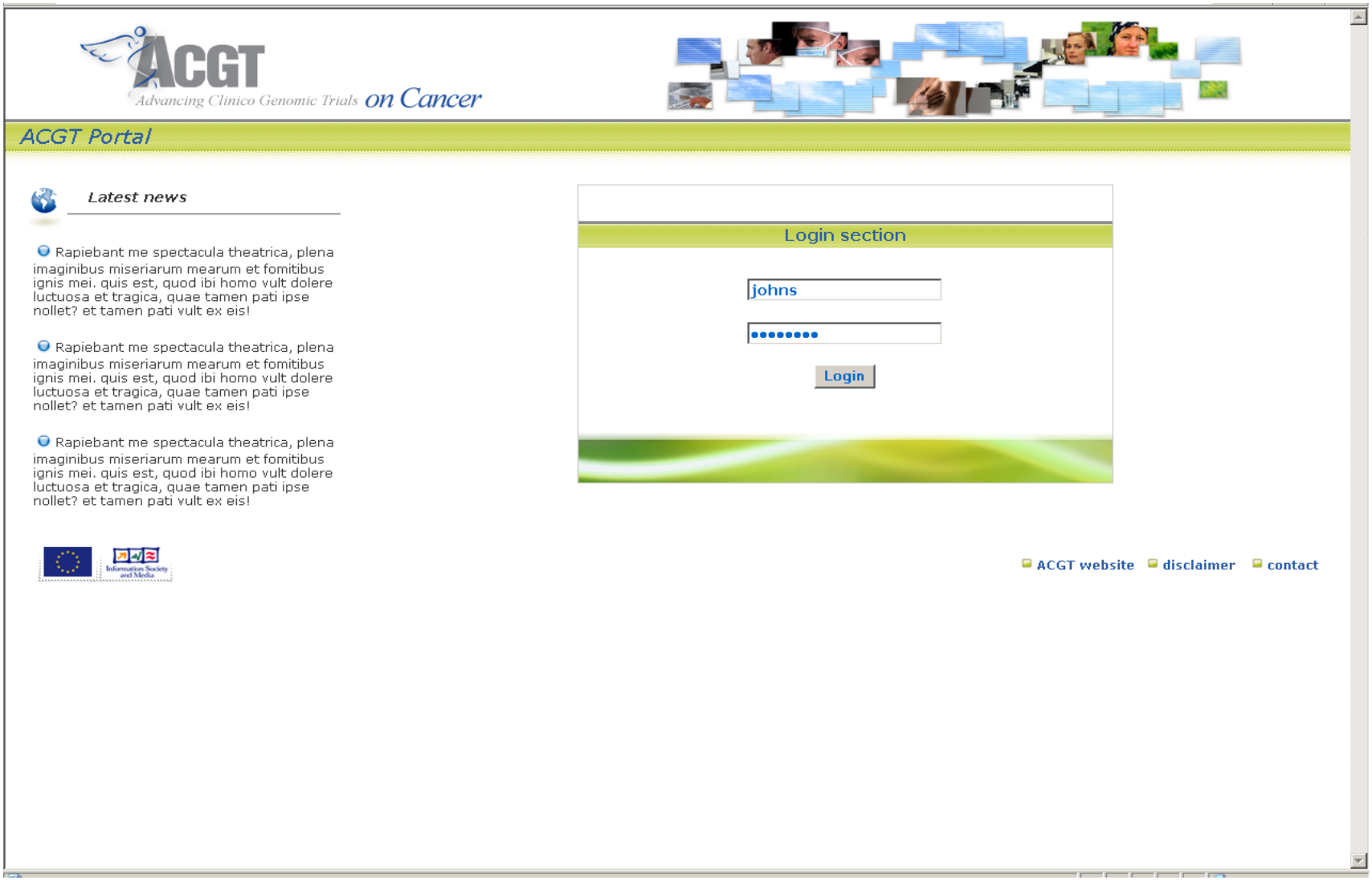

Figure 4: ACGT Portal – Login page prototype

In the visual prototype of the homepage, large blocks are displayed in two third of the page and aligned to the right. Small blocks are displayed in one third of the page and aligned to the left.

For a given user, each block of content can be maximized (displayed in the large shape), minimized (displayed in the small shape) or hidden (not displayed in a certain page or in all pages).

The map of blocks is initially configured in the standard profile of a given role. After signing-in, a user can customize the display of the content blocks. The customization of the content blocks presentation can be done:

- From each page, by using the buttons available in the block bar.
- From the My Profile page, which is accessible to each user.

A visual prototype of the My Profile page is presented in **Figure 6:** ACGT Portal – User's profile page prototype.

In each of these configuration variants, also the position of each block in the page can be decided. In order to keep the complexity of these operations to a low level, only vertical ordering of blocks (in the two coloumns, left/right) would be allowed.

In the default configuration, the blocks can be ordered using general criteria, like frequence of access or average position in the profiles of all users of the same type, etc.

6.2 Search methods

Finding a resource in a portal can be done either by browsing the structure of the portal or by direct searching with different criteria.

6.2.1 Browsing

The content of the portal can be organized on:

- Content categories.
- Content types.

A content category is a folder that may contain items of different types, which are related to each other by some logical properties. An item can simultaneously belong to several categories, but at least one.

In complex systems, like the ACGT Grid Infrastructure, the organization of content in categories is essential and should be done on a methodological base. The ACGT Master Ontology that will be developed within the WP 7 of the project is the main candidate for representing the structure of the ACGT Portal on which the content can be distributed.

However, parts of the ACGT Master Ontology can be irrelevant for some user roles in the ACGT Portal. The best approach is to adapt the structure of the ACGT Portal to each role in the portal.

Figure 5: ACGT Portal – User's homepage prototype

The first criteria to be used in drawing the map of the ACGT Portal adapted to a certain role in the portal is to display only ontological categories that contain resources that are accessible in some way to the user playing that role in the portal.

The second criteria to be used in the organization of content personalized by the roles in the portal would be to generate different views of the ACGT Master Ontology based on the way(s) in which each user accesses the content in the portal. This kind of analysis cannot be done otherwise than tracking the users activity in the portal (see next section). This analysis and the implementation of its results shall be performed on the ACGT Portal prototype that will be deployed by the end of Month 12 of the project.

6.2.2 Searching

Regardless how good is the organization of the portal, many users will prefer to look for the resources in the portal, not by browsing the structure, but by searching directly the desired resource.

Efficient search methods are usually simple (free questions if possible) and transparent to the user. Developing advanced search methods like ontological search or data mining is a desirable orientation. However, the risk that advanced search methods will become too sophisticated and consequently inaccessible to most of the users should be avoided.

Search methods will be implemented in the ACGT Portal as they are developed by the corresponding WPs which are related to service development.

6.3 User activity repository

Users activity in the portal is an important source of information. Keeping the history of the users activity in the ACGT Portal is equally important for performing statistics and for providing users with a repository for the last actions in the portal.

The tracking of the users activity in the portal shall be done by respecting all legal and ethical norms.

6.3.1 Statistics

Statistics are very important for representatives of different stakeholders (cancer organization representatives, regulatory agencies representatives).

Statistics have to express different indicators regarding the activity of the users in the ACGT Portal:

- Number of visits per each page.
- Number of visits per each type of item.
- Different types of correlations, like:
    - Users with the same behaviour (searching the same items, performing the same actions).

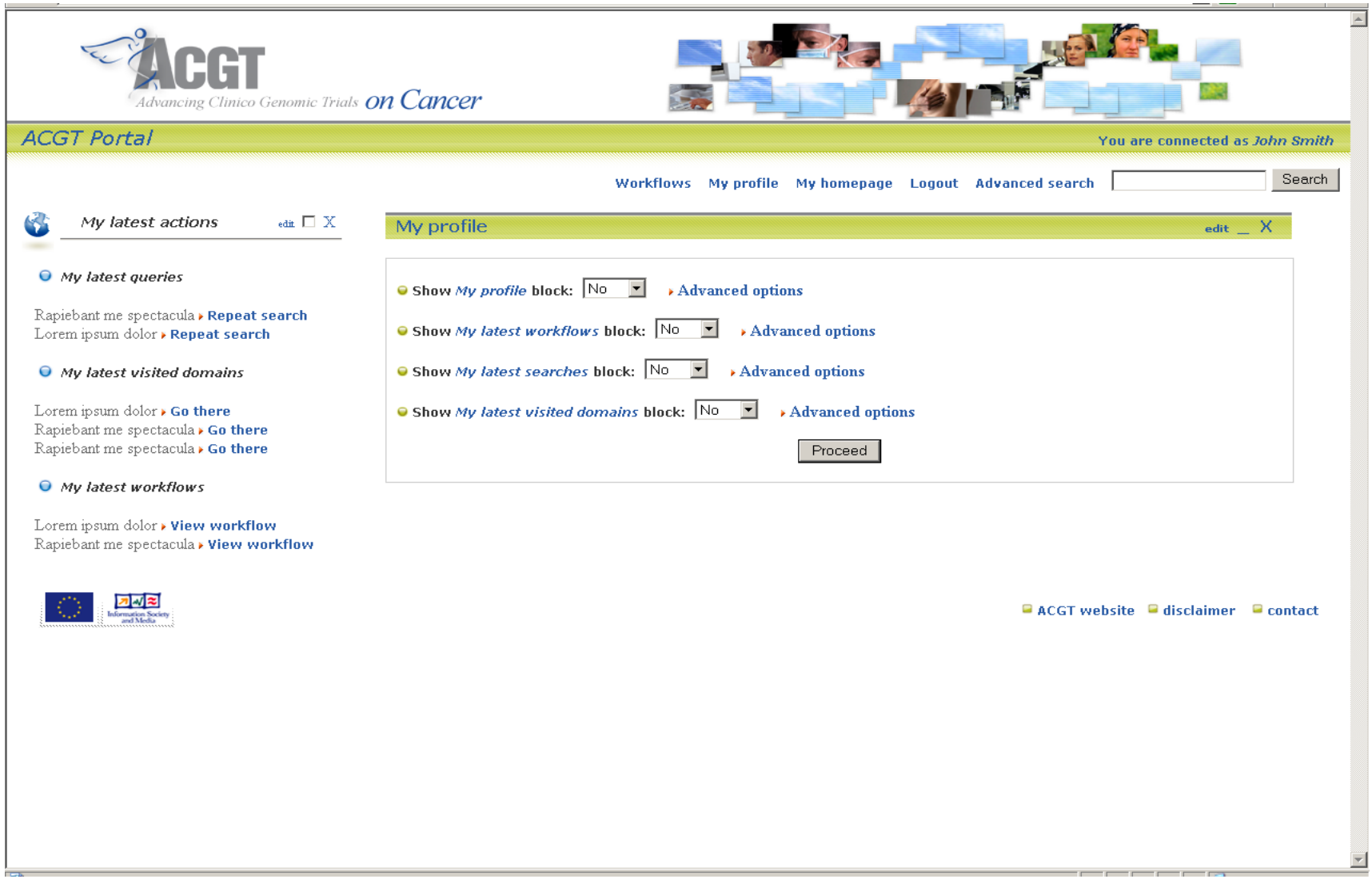

Figure 6: ACGT Portal – User's profile page prototype

- Similarity between items.
- Searching paths.
- Browsing paths.

Statistics can contribute to the better understanding of users preferencies and to map the users behaviour in the portal.

Such an analysis will eventually lead to the reorganization of the content in the ACGT Portal based on the users preferencies and behaviour.

6.3.2 User history

The history of the user activity in the portal is an important feature of the ACGT Portal that can contribute to the fidelization of users.

User activities in the portal are grouped on types of activities, like:

- History of workflows.
- History of searches.
- History of messages.

This information will be displayed in the homepage of the user (see **Figure 5:** ACGT Portal – User's homepage prototype) and in other pages where these blocks of content are selected (see **Figure 7:** ACGT Portal – Workflows page prototype).

The history of users' activity would allow users to:

- Repeat the actions in the portal, possibly with different input data.
- Launch automatic actions without being forced to stay connected to the portal and retrieve the results of these actions.

For example, the user would be able to perform the following sequence of steps:

- Select an workflow.
- Upload the input data.
- Configure the workflow:
  - Select databases where searches to be performed.
  - Select tools to perform data in different steps of the workflow.
- Launch the workflow.
- Log-out.
- Log-in (after some time).

Figure 7: ACGT Portal – Workflows page prototype

- Retrieve the (intermmediate or final) results.
- Change some input data or configuration variables.
- Re-run the workflow and get new results.

A visual prototype of workflow presentation page is given in **Figure 8:** ACGT Portal – Specific workflow page prototype.

Another important application of this function of the ACGT Portal would be the possibility of sharing data and actions in restricted groups of users (see Section 6.6 Collaborative work).

Due to space restrictions, this function of the portal can be limited to a number of past actions. However, special arrangements of distributing the hosting of these historical databases on several locations and/or providers would allow rather generous limitations for the number of past actions, at least for selected users. For example, an organization can host the historical data of its own members.

6.4 User online assistance

User assistance is an important component of the ACGT Portal. The principle governing this functionality of the portal is that the user support should be provided in the place and at the time when the user needs it.

6.4.1 Training materials

Training materials that will be offered through the ACGT Portal are:

- Courses.
- Tutorials.
- Research reports.
- Video materials.

Training materials will refer both to the use of the ACGT Portal facilities and to the scientific matters, as performing medical investigations, participating in trials, etc.

6.4.2 Contextual help

Contextual help will be provided in each page in two forms:

- Small boxes of information attached to different actions that can be performed in a page.
- Links from the scientific words in the texts to the explanations of these words in the ACGT Master Ontology.

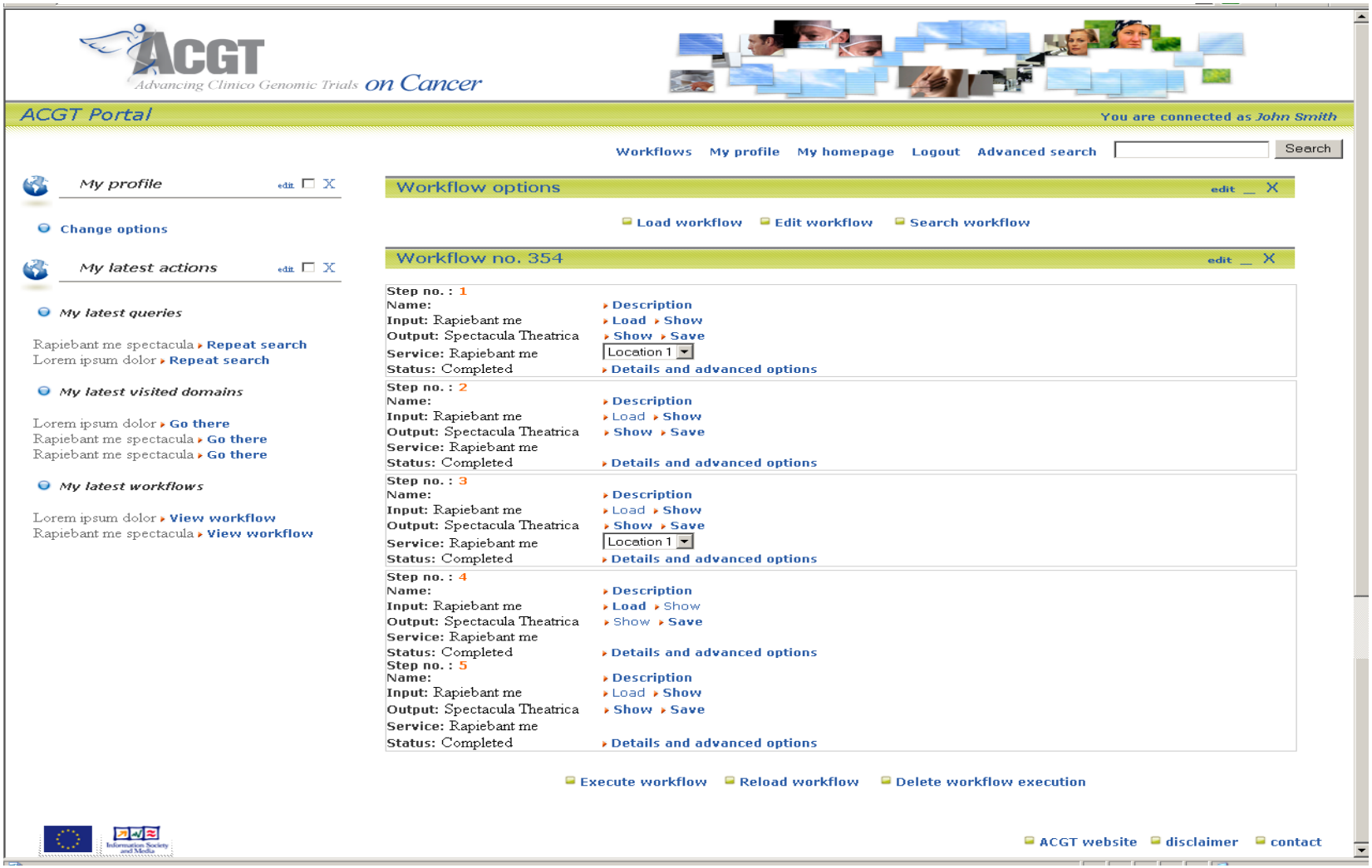

Figure 8: ACGT Portal – Specific workflow page prototype

6.4.3 Interactive demos

The implementation of interactive demos in the ACGT Portal is another user assistance component. Interactive demos are particularly suitable for assisting the execution of workflows. The interactive demo of a workflow will provide the user with the demonstration of:

- How to upload correctly the input data.
- What variance is allowed in the type and quantity of input data.
- How the workflow can be configured.
- What differences can be made by configuring the workflow in a certain way.
- Which are the predictable results for a certain configuration and set of input data.
- How results can be interpreted.
- How the run of the workflow can be modified in order to vary the results.

The development of interactive demos will be done using a methodology that will be defined and tested in the Task 14.4 of the project.

6.5 Detailed scheme of privilegies and roles

The definition of a complet scheme of privilegies and roles is essential for the management of the content.

The detailed scheme of privilegies and roles of the ACGT Portal is presented in Chapter 5 of this document.

6.6 Data and resource security

Data and resource security is concerned both with the (non)-access to data concerning patients and to data and resources that are in a research stage.

6.6.1 Patient data

Sensitive data, like pseudonimized data on clinical cases will be manipulated through the ACGT Portal.

In this respect the security implemented in the ACGT Portal should be consistent with the general security policy of the ACGT infrastructure.

Sensitive data should be available only to the specific users and only in the form that is established in the general policy of the ACGT infrastructure. The trust and security aspects of the ACGT infrastructure are approached within the WP 11.

6.6.2 Private data

On the other hand, data or resources produced in clinical trials that is not yet tested and confirmed will be manipulated through the portal within restricted groups of clinicians and researchers.

The access to the data and resources should be 100% modelled upon the scheme of privilegies and roles.

Security policies should be implemented through different methods, focused on:

- Groups of users.
- Individual users.
- Portlets.
- Data.
- Resources.

6.7 Collaborative work

The collaborative features provided by the ACGT Portal will be essential for the development of the scientific community around the ACGT Infrastructure.

6.7.1 Groups of users

The main collaboration unit will be the group of users, leaded by an editor and grouping users that are performing a collaborative activity (for example, a trial).

The group will have its own homepage, where the editor will be allowed to customize the presentation of the content. Items published under the group scheme will be either private or public according to the approval of the editor.

The members of a group will be allowed not only to share content (news, documentation, scientific papers, messages, etc) but also to share the run of workflows or other applications. This will be possible through the group activity repository that will keep the history of the actions performed by the members of the group

6.7.2 Sharing content

Multiple forms of collaboration will be defined in the ACGT Portal. These forms of collaboration include:

- File sharing
- Application sharing
- Discussion forum

PART 4

Technical specifications

7 Platform description

In traditional web application development, very few libraries exist to make portal development easy. In general, many homegrown and vertical solutions exist and very little code is shared or reused. Finally with the emergence of some critical new web technologies, web application development is focusing more on reusable solutions and software.

The GridSphere portal framework provides a standards based portal for the easy development of modular web components, called 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.

This chapter discusses briefly the the GridSphere portal architecture including the layout engine, support for two portlet API implementations and the portlet services model. This chapter is based on the paper [NOV2003] J. Novotny, M. Russell, O. Wehrens, GridSphere: An Advanced Portal Framework. For more information, the reader is invited to visit the Gridsphere website: http://www.gridsphere.org/.

7.1 Introduction

Although many definitions exist for the term web portal, most definitions classify a portal as a gateway web site that offers a set of services. The services provided may range from weather information to stock quotes to the ability to search for other sites and many other features. Web portals may also offer personalization and customization features that provide users with a unique, individualized web experience.

Examples of generalized portals include Excite, Yahoo, America Online, or Microsoft Network . Niche portals tend to offer specialized content suitable for a particluar audience. CNN.com is an example of a news portal or the increasingly popular Friendster portal supports building personal connections and communities of people with common interests.

In general, many different technologies may be used to construct a portal. A portal can be delivered using anything from a loosely couple collection of static web pages to using a full-blown application server. An application server is viewed as a part of a three-tier application consisting of the following components as shown in the diagram:

1. The first tier is a user's web browser, generally on their desktop but may also be a mobile device or PDA.

2. The second tier web performs business logic and generally combines or work with a web server that can dispatch requests via HTTP from the first-tier web browser.

3. The third tier, or back-end, are resources such as databases, information servers or compute resources.

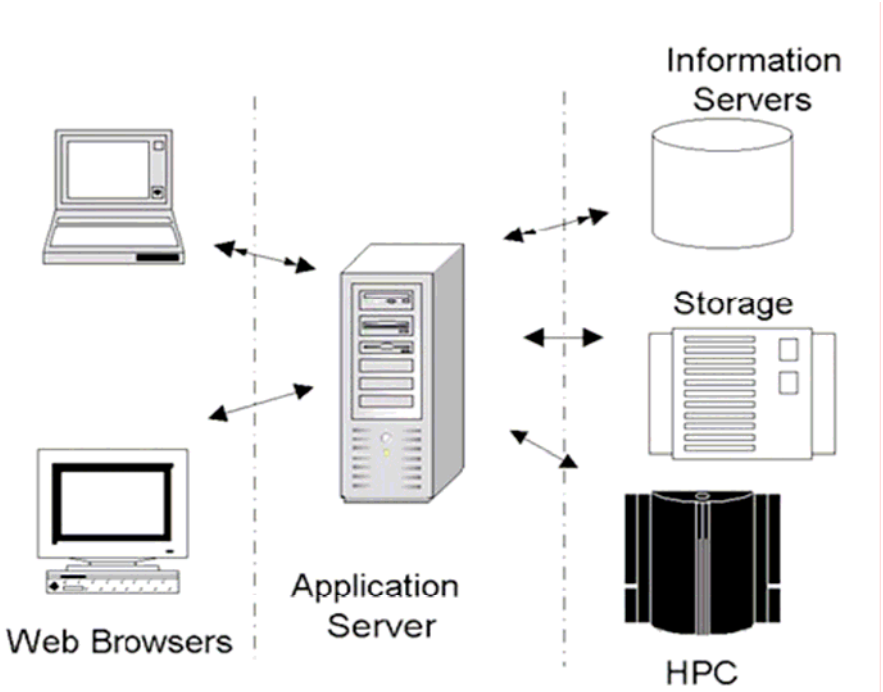

Figure 9: Classic 3 tier application architecture

A wide variety of software packages exist that perform the duties of the application server. All in one solutions such as IBM's WebSphere, Oracle i9AS Server, Sun Application Server or BEA WebLogic provide monolithic solutions that may be relatively easy to set up and configure "out of the box".

Many open source solutions are based on using the popular Apache web server, the Jakarta Tomcat[WEBTOM] Servlet container, in combination with various web application frameworks.

The vast majority of application server solutions are currently JavaTMbased over any other implementation technology for the reason that server side Java has gained tremendous popularity as a language that provides greater support for the development of component based architectures and can more readily separate presentation from logic. According to a Jupiter report [JUP2003] of the most widely used application servers, over 95% of the portal market is dominated by Java based technologies.

While a portal can be constructed simply by using a web server and a collection of static HTML pages, most portals capable of any real functionality especially personalization, security and transactions, require an architecture for the development of components that can add new functionality and provide reusability.

Web application frameworks have emerged as a collection of tools and best practices to make the development of new components easier. With the last years the number of application frameworks has ballooned.

Currently, a list of the most popular web frameworks includes Struts[WEBSTR], Tapestry[WEBTAP], WebWork[WEBWOR], Turbine[WEBTUR], Barracuda[WEBBAR], Expresso[WEBEXP], Spring[WEBSPR] and numerous others. To the first approximation all frameworks have one goal in common which is to provide a mechanism for the separation of logic and presentation. Quite clearly this separation provides far greater flexibility in the

features and functionality that can be provided by the application server and how it gets displayed and possibly customized to end users. In general, all frameworks make use of a very common and popular design pattern for this separation called, the Model View Controller (MVC) pattern. In the MVC approach, the model is comprised of the data that is to be displayed and is generally encapsulated in terms of some generic abstract data type. The view refers to the component responsible for the display of the model and can be specially tailored or customized. The controller has the responsibility of dispatching requests from the user to the appropriate set of components that construct the model and presentation.

Each of these three aspects represent orthogonal functionality that can be varied independently resulting in a powerful design paradigm. As will be discussed in the next sections, the portlet model also encapsulates MVC design and can be complemented with existing web application frameworks. In addition the portlet model has been ratified through a formal specification unlike existing web application frameworks giving it a broader appeal.

7.2 The Portlet Concept

Last October 2003, the portal market was shaken up when a new API, known as Java Specification Request (JSR) 168[WEBJSP] Portlet API, was finally approved by most of the major portal application server vendors including IBM, Sun, Oracle, Plumtree, BEA, Vignette, SAP, and ATG runder the Java Community Process (JCP).

The Portlet API represents an attempt at creating a well defined interface for the development of web components that can be shared and exchanged among portal vendors to provide enhanced functionality. For years many vendors had been developing their own proprietary models for the creation of new components that would work with their specific portals but could not be reused in another vendor's portal.

The Portlet API represents a convergence of views that just as operating system lock-in prevents the sharing of applications between say MicrosoftTM, AppleTMand Linux, various third party software providers needed to tailor their specific set of web functionality for individual portals resulting in far greater development effort, lack of any real reusability and overall maintenance headaches. It is a fairly good guess that while portal vendors will continue to vie for customers by providing overall greater ease of use, configuration and bundled software with their portal offerings, the shift has now gone to producing portlet applications that provide "add-on" value for specialized customer bases. Independent software vendors (ISV's) can now concentrate on delivering functionality in particular focused areas and program to a common API.

Portlets are poised to revolutionize web application development by providing a standardized, well formulated model supporting MVC design. The existence of so many other web development frameworks proves how necessary this is in the real world, but none until have now have been standardized and received so much vendor support.

Portlets[WEBIBM] define web application components with a well defined set of lifecycle methods much like the Java Servlet API[WEBSER] as well as providing a visual interface to a content or service provider. From a technical perspective, portlets represent modular, reusable software components that may be developed independently of the general portal architecture and offers a specific set of operations. For instance, portlets may provide users with an updated list of stock quotes or content from a news feed. Visually portlets appear as

”mini-windows” within a portal page and have a title bar providing options for setting the portlet mode and window state.

Just as a servlet container provides a runtime environment for the loading and management of Java servlets, a portlet container performs the necessary initialization and management of portlets. The portlet container acts as an intelligent dispatcher to forward client browser requests to the appropriate portlet during a request response cycle. It is the responsibility of the portal and not the portlet container to appropriately render the markup of portlets onto a single portal page.

The next sections will discuss the architecture of the GridSphere portal consisting primarily of the portlet container and the portal layout engine.

7.3 GridSphere Framework Overview

GridSphere provides a portlet implementation, a portlet container and a collection of core services and portlets as shown in the diagram:

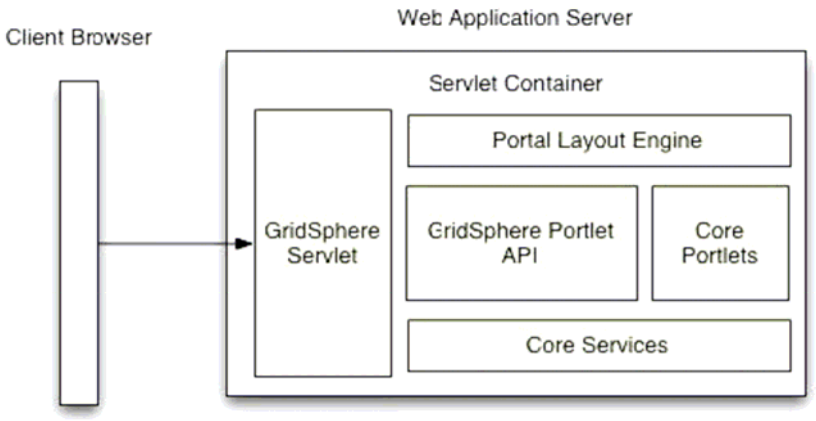

Figure 10 GridSphere Portal Architecture

The architecture diagram shows the primary components that combined with the servlet container comprise the web application server. A request to the portal triggered by a user’s web browser invokes the GridSphere servlet which acts as a controlling dispatcher to the layout engine responsible for rendering suitable output to the user’s browser.

Both the GridSphere servlet and layout engine make use of core services, including the portlet registry, to invoke the appropriate set of portlets in a users portal page.

The following subsections discuss each of these components of the architecture in greater detail.

7.3.1 The Layout Engine

Layouts in the portal are defined as XML descriptor files as shown in the following example snippet which displays the portal page banner:

```
<page-layout theme="xp" title="GridSphere Portal">

<portlet-header>
 <table-layout style="header">
  <row-layout>
   <column-layout width="70%">
    <portlet-content include="/html/pagehead.html"/>
   </column-layout>
   <column-layout width="15%">
    <portlet-frame transparent="true" label="locale">
     <portlet-class>LocalePortlet.1</portlet-class>
    </portlet-frame>
   </column-layout>
   <column-layout width="15%">
    <portlet-frame transparent="true" label="logout">
     <portlet-class>LogoutPortlet.1</portlet-class>
    </portlet-frame>
   </column-layout>
  </row-layout>
 </table-layout>
</portlet-header>
...
```

Figure 11: Layout descriptor

The nested structure of the layout descriptor provides a very handy approach to modeling a templated layout that can be easily modified that acts as an abstraction layer above the underlying display technology, generally HTML. The Composite design pattern is employed for the representation of layout components within a nested tree-like structure. Each layout component defined in Grid-Sphere e.g. PortletFrame, PortletTabbedPane, PortletTab, PortetContent, etc itself adheres to a PortletComponent interface which defines a set of lifecycle methods that closely maps to the lifecycle methods of portlets. The following UML sequence diagram shows how the actual portal page rendering is accomplished:

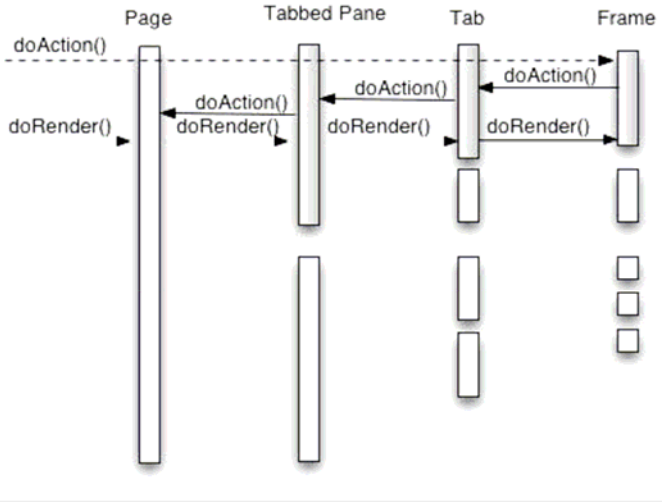

Figure 12: Layout Engine Sequence

The diagram is simplified by showing a page consisting of a tabbed pane which itself contains a few tabs with frames within the tabs. The frames represent the visual appearance

of the portlet and are responsible for obtaining a portlet's markup for display in the portal page. In typical GridSphere usage, the portal page uses a double tabbed pane approach, where each tab itself contains an additional tabbed pane with tabs and frames for greater navigability like that supported by many portal sites including the Amazon, and Apple websites.

As shown, an event caused by clicking on a layout component performs an action in that component which is then ”bubbled” upward to to the root of the layout tree. Afterwards, the component tree is rendered appropriately. Notice that not every component in the tree is rendered. For instance only the visible tab will continue to propagate render events to its children components.

The Layout Engine offers great flexibility and is entirely customizable. New components can simply be added by creating an XML descriptor element and a corresponding component class that implements the PortletComponent interface. This model was extended by other projects wishing to build specialized visual components for rendering iFrames, XHTML or that provide additional layouts.

It should be stated that no other portal implementation supports such a generic model for the creation and usage of layout components.

Most usually offer a small handful of simplistic web page templates that can be edited, but none actually allow a developer to create entirely new layout components that may be presented in the portal.

7.3.2 The GridSphere Portlet Model

When GridSphere was first under development, the ratified JSR 168 Portlet API did not exist and so we had two options. One, we could have chosen to build off of the open source Jakarta Jetspeed project[WEBJET]. In fact, at the time, Jetspeed was the most highly evolved open source portal offering a first attempt at developing a Portlet API.

However, after further evaluation[NOV2002], it was determined that Jetspeed 1 had many dependencies on other large stacks of software such as the Jakarta Turbine project that were quickly changing. In addition, the Portlet API offered was not quite robust enough to support the easy development and packaging of external, or ”third-party”, portlets.

However, upon looking at the IBM WebSphere 4.1 Portlet API we found a richer, better documented API that itself had been based upon Jetspeed in earlier versions of WebSphere. Believing the finalized API to be somewhat similar to the WebSphere rAPI, we chose to base our implementation using the IBM WebSphere 4.1 Javadocs[WEBSPH] and Developer's Guide[HES2002] as a starting point.

For the first year of development, we implemented the WebSphere Portlet API and based our core portlets provided by GridSphere on this model. One of the benefits of the IBMrPortlet API was a clearer packaging model that allowed for easy integration of thirdparty portlets as Web Archive, or WAR files, in the same way the Java Servlet API specifies the packaging of web applications as WAR files.

In the WebSphere portlet model, there are no interfaces defining a portlet. Instead a portlet is represented as a hierarchy of abstract and base classes as shown on the left. A portlet developer would subclass from the AbstractPortlet base class to develop a customized portlet, MyPortlet in the diagram.

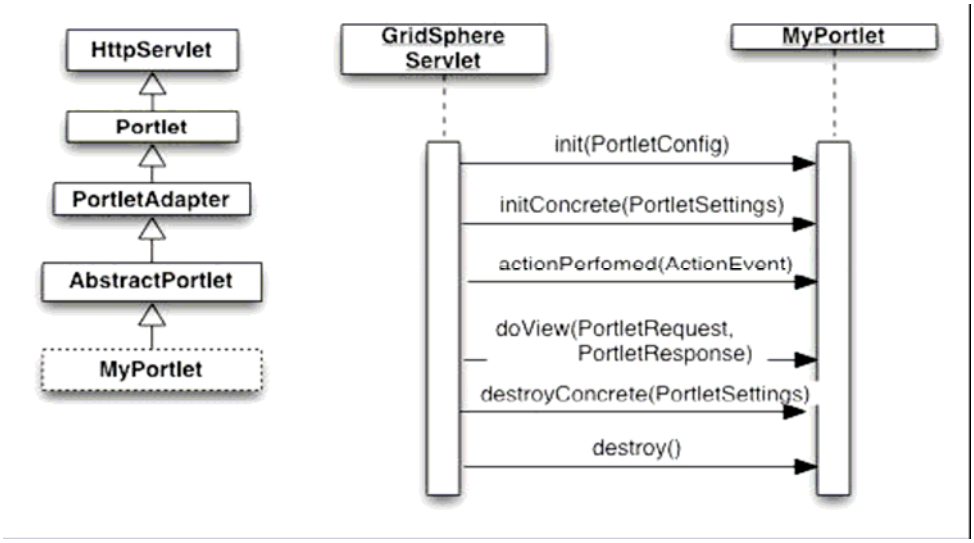

Figure 13: IBM WebSphere portlet hierarchy and sequence

The WebSphere rportlet model makes a distinction between Application Portlets and Concrete Portlets where a concrete portlet is a parametrized instance of an application portlet. For instance a stock quote portlet can have multiple concrete instances that are parametrized with different stock quote services from which the quotes are obtained.

As shown in the sequence diagram, an application portlet and its corresponding concrete portlets are initialized during init() and initConcrete() method invocations respectively. Similarly, portlets are destroyed via the destroyConcrete() and destroy() methods. These methods generally occur only once during a portlets lifecycle, while the actionPerformed() and doView methods may occur during every client request potentially. The actionPerfomed is called only when an action occurs in the portlets, usually as a form submission, button click or portlet hyperlink being invoked. For simplicity only the doView() method is shown in the sequence, although portlets in allow for three additional portlet modes: Edit, Configure and Help.

A portlet developer may implement the doEdit(), doConfigure() or doHelp() methods to support these modes when the mode icon is selected from the portlet title bar. The appropriate render method is invoked for every portlet that is displayed on the page, although the portal may choose to cache rendered output for redisplay.

Just as the portal layout is defined by a layout descriptor, a portlet descriptor file, portlet.xml, defines the portlet configuration information and provides the portlet container the necessary details for instantiation of the portlets.

Both the PortletConfig and PortletSettings objects used during portlet initialization contain configuration information obtained from the portlet deployment descriptor.

The JSR 168 Portlet API is relatively similar to the WebSphere Portlet API. However, in JSR 168, a portlet is defined by a Portlet interface which is implemented by the GenericPortlet base class as shown in the diagram:

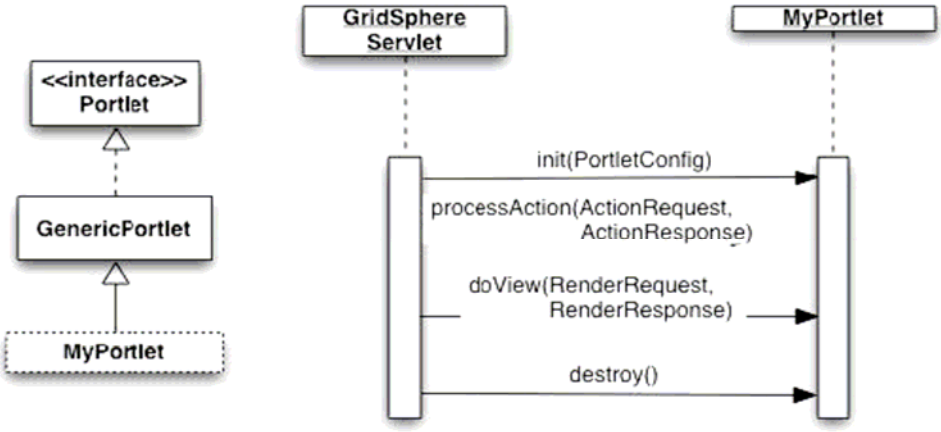

Figure 14: JSR 168 portlet hierarchy and sequence

A portlet developer would subclass from GenericPortlet as shown to produce MyPortlet as an example. The lifecycle interfaces provided in the JSR 168 Portlet API are similar to the WebSphere API, however there is no notion of concrete and application portlets. Only one portlet exists and is initialized. Instead of ActionPerformed(), the action method that is invoked when a portlet action occurs is replaced by the processAction() method.

In practice, it is not too difficult to refactor WebSphere portlets to be JSR compliant, although the WebShere API provides a few additional interfaces for portlet messaging and better localization support.

After JSR 168 had been finalized and ratified we faced the design decision of how to support the JSR portlet API and handle the refactoring of existing portlets that would be necessary. Upon further reflection, a plan was devised that would continue to support the existing WebSphere API and offer support for JSR 168. The merits of doing this are the following:

1. Allows developers that have already developed WebSphere rportlets to keep existing portlets without having to refactor to JSR portlet API.

2. Continues to supportWebSphere rAPI compliance allowing WebSphere users to migrate to GridSphere as an open source portal solution.

3. Does not force existing GridSphere framework code to be refactored which would destabilize GridSphere for some period of time

4. Modularity allows JSR portlet API implementation to be developed on top of existing functionality and requires only minimal modifications to the existing framework.

The following architecture diagram shows how support for both WebSphere and JSR 168 portlet API's is provided. The portlet registry acts as a repository for portlets that is used by the GridSphere servlet controller and the layout engine.

In the case of a WebSphere portlet application, the portlets are registered in the portlet registry upon startup. To the GridSphere framework, a JSR portlet application looks just like a WebSphere application with one special portlet, the PortletServlet .

The PortletServlet performs the task of loading and registering all JSR portlets in the application, such that the portlet registry contains both WebSphere and JSR 168 portlets that can be invoked by the GridSphere servlet and the layout engine.

In addition, the Grid-Sphere portlet container includes portlet API implementations of both the WebSphere rand JSR 168 API's, such that portlet development for both models is supported.

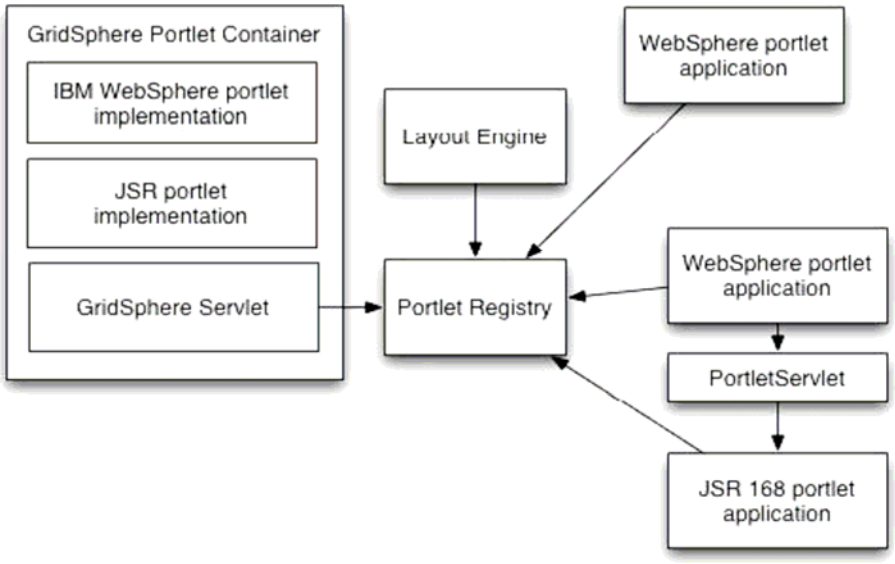

Figure 15: GridSphere support for WebSphere and JSR 168 portlet API's

The following sequence diagram shows how the container is responsible for invoking JSR portlet lifecycle methods via the special PortletServlet gateway.

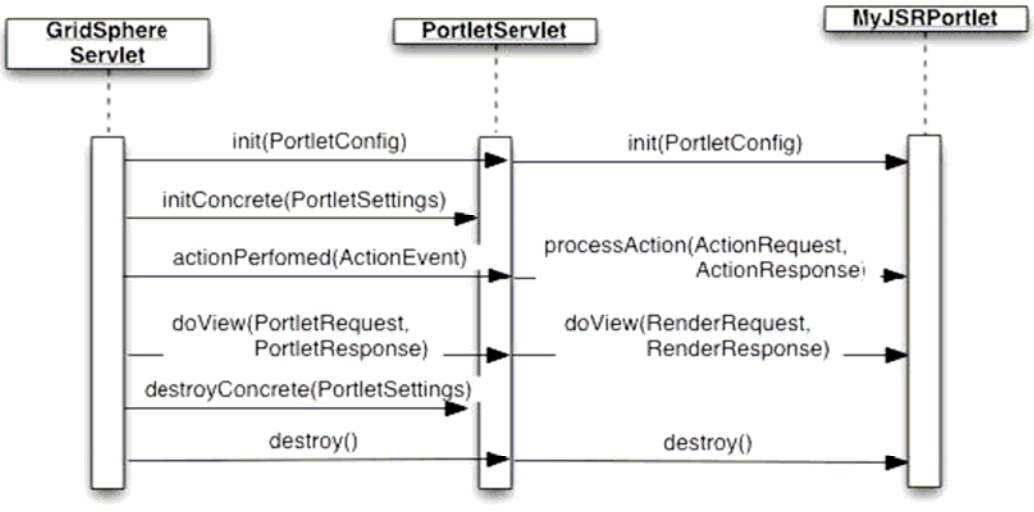

Figure 16: Sequence diagram capturing both portlet API's

Generally the PortletServlet is responsible for translating WebSphere portlet lifecycle methods into JSR portlet lifecycle methods. Fortunately, the WebSphere model contains a superset of the JSR portlet methods that translate quite easily. In the case of an initConcrete() or destroyConcrete() method call, no JSR portlet invocation is performed since

that concept does not exist. The PortletServlet also does the work of translating the WebSphere portlet API core objects such as PortletRequest or PortletResponse into the RenderRequest or RenderResponse objects as required by the JSR 168 portlet model.

7.3.3 Portlet Services Framework

The portlet services framework defines an architecture for the development of portlet services that provide functionality to portlets. Portlet services are designed to individuate the functions provided by portlets from the services with which they need to interact with. Portlet services provide an encapsulation of reusable business logic that may be reused by one or more portlets. In GridSphere, services are used to manage everything from layout preferences, user profiles, user access control as well as the portlet registry discussed previously.

Portlet Services API

The IBM WebSphere API provides a set of interfaces for defining portlet services and allows developers to get portlet service instances via the PortletContext object.

Although the JSR portlet API makes no mention of portlet services, Gridsphere can provide access to services by using its own ”enhanced” portlet which subclasses from the base GenericPortlet.

A portlet service is created from a PortletServiceFactory. A PortletService defines a blank interface that is enhanced by the PortletServiceProvider interface which defines a lifecycle consisting of init and destroy methods used when the service is initialized and shutdown.

A portlet service configuration file is used to provide class information and initialization parameters that is accessed via the PortletServiceConfig object. The PortletServiceFactory is also responsible for initializing and destroying services. A portlet service UML class diagram is shown in the below figure.

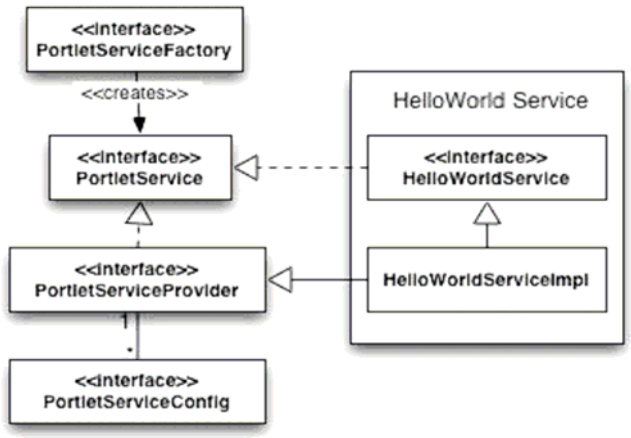

Figure 17: Portlet services UML class diagram

Like the portlet or layout descriptor, the portlet services descriptor is also expressed as an XML file. An example entry is shown in the next figure.

```
<service>
 <name>Login Service</name>
 <description>Provides Login Capabilities</description>
 <interface>LoginService</interface>
 <implementation>LoginServiceImpl</implementation>
</service>
```

Figure 18: Portlet services descriptor

7.4 GridSphere Core Portlets

The GridSphere framework provides a core set of portlets that offer the basic functionality required for the portal to be usable. The following list itemizes the core set of portlets provided:

- Login - Allows users to login based on a name and password
- Logout - Logs a user out of the portal
- Locale Selection - A user can select from English, French, German, Italian, Hungarian, Czech and Polish
- Account Request - Provides interface for a new portal user to request an account and optionally choose groups to join
- Account Management - Provides the Super user and Admin users the ability to assign/revoke
- users roles
- User Management - Provides the Super user the ability to approve/deny account requests, and Admin users the ability to approve/deny group requests
- User Profile - Provides the ability for users to edit personal information including name, email address and select portlet applications to join
- Layout Configuration - Users can configure their layout including placement of tabs and portlets within tabs
- Portlet Subscription - Provides the ability for users to add and remove portlets from their workspace
- Local File Manager - Provides a virtual filesystem allowing users to edit, upload and download files to the portal
- Notepad - Users can create, edit, delete and search notes
- Text Messaging - Users can text message other users using instant messenger

7.5 Conclusion

The Portlet API has garnered a lot of support within the web application server community and looks to be a viable model for the sharing and reuse of portal components. The GridSphere portal framework provides both an implementation of the WebSphere rand JSR 168 Portlet APIs as well as a full-fledged portal and portlet container. The collection of core portlets bundled with GridSphere offer the minimum functionality required of a portal including user and group management, user customization and user profile management and more.

The main reasons for choosing to build up the ACGT Portal on the Gridsphere Portal are:

- Standards compliance
- Open-source platform
- References in the domain of biomedical portals.

7.6 References

[NOV2003] J. Novotny, M. Russell, O. Wehrens, GridSphere: An Advanced Portal Framework, 2003

[JUP2003] Jupiter Application Server Report
HTTP://WWW.NETWORKWORLD.COM/NEWS/2003/0214STUDYORACL.HTML

[WEBSTR] Jakarta Struts framework
HTTP://JAKARTA.APACHE.ORG/STRUTS/INDEX.HTML

[WEBTAP] Jakarta Tapestry framework
HTTP://JAKARTA.APACHE.ORG/TAPESTRY/INDEX.HTML

[WEBWOR] WebWork
HTTP://WWW.OPENSYMPHONY.COM/WEBWORK/

[WEBTUR] Jakarta Turbine framework
HTTP://JAKARTA.APACHE.ORG/TURBINE/

[WEBBAR] Barracuda
HTTP://BARRACUDAMVC.ORG/BARRACUDA/INDEX.HTML

[WEBEXP] Expresso framework
HTTP://WWW.JCORPORATE.COM/

[WEBSPR] Spring framework
HTTP://WWW.SPRINGFRAMEWORK.ORG/

[WEBSER] Java Servlet 2.3 and Java Server Pages 1.2 Specifications
HTTP://JAVA.SUN.COM/PRODUCTS/SERVLETS/

[WEBTOM] The Jakarta Tomcat Project
HTTP://JAKARTA.APACHE.ORG/TOMCAT/

[WEBJET] The Jakarta Jetspeed Project
HTTP://JAKARTA.APACHE.ORG/JETSPEED/

[WEBIBM] What is a Portlet?
HTTP://WWW-3.IBM.COM/SOFTWARE/WEBSERVERS/PORTAL/PORTLET.HTML

[WEBJSR] JSR 168: Portlet Specification
HTTP://WWW.JCP.ORG/JSR/DETAIL/168.JSP

[HES2002] Hesmer S., Fischer P., Buckner T., Schuster I., Portlet Development Guide April 2, 2002;

[WEBSPH] WebSphere rPortlet API Javadocs
HTTP://WWW-106.IBM.COM/DEVELOPERWORKS/WEBSPHERE/ZONES/PORTAL/PORTLET/4.2API/WPS/

[NOV2002] J. Novotny, I. Kelley, M. Russell, O. Wehrens, Jetspeed Evaluation, May, 2002