Developing a European Biomedical GRID for post-genomic research on Cancer

M. Tsiknakis, D. Kafetzopoulos, G. Potamias, A. Analyti, K. Marias, S. Sfakianakis

***Abstract*—This paper presents the needs and requirements that led to the formation of the ACGT (Advancing Clinico Genomic Trials on Cancer) integrated project, its vision and methodological approaches. The ultimate objective of the ACGT project is the development of a European biomedical grid for cancer research, based on the principles of open access and open source, enhanced by a set of interoperable tools and services which will facilitate the seamless and secure access to and analysis of multi-level clinico-genomic data, enriched with high-performing knowledge discovery operations and services.**

By doing so, it is expected that the influence of genetic variation in oncogenesis will be revealed, the molecular classification of cancer and the development of individualised therapies will be promoted, and finally the in-silico tumour growth and therapy response will be realistically and reliably modelled.

The scenario-based requirements engineering methodology adopted by the project is presented together with indicative post-genomic such scenarios. Subsequently, the main technological and research challenges of the project are presented together with the methodological approaches adopted for addressing them.

I. Introduction

This is a critical time in the history of cancer research as recent advances in methods and technologies have resulted in an explosion of information and knowledge about cancer and its treatment. As a result, our ability to characterize and understand the various forms of cancer is growing exponentially, and cancer therapy is changing dramatically. Today, the application of novel technologies from proteomics and functional genomics to the study of cancer is slowly shifting to the analysis of clinically relevant samples such as fresh biopsy specimens and fluids, as the ultimate aim of translational research is to bring basic discoveries closer to the bedside.

The implementation of discovery driven translational research, however, will not only require co-ordination of basic research activities, facilities and infrastructures, but also the creation of an integrated and multidisciplinary environment with the participation of dedicated teams of clinicians, oncologists, pathologists, epidemiologists, molecular biologists, as well as a variety of disciplines from the domain of information technology.

Today, 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 to the research community at large, presents 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 capitalize on these opportunities [1]. Much of the genomic data of clinical relevance generated so far are in a format that is inappropriate for diagnostic testing. Very large epidemiological population samples followed prospectively (over a period of years) and characterized for their biomarker and genetic variation will be necessary to demonstrate the clinical usefulness of these tools.

Up to now, the lack of a common infrastructure has prevented clinical research institutions from mining and analyzing disparate data sources. This inability to share technologies and data developed by different cancer research institutions can therefore severely hamper the research process. Similarly, the lack of a unifying architecture is proving 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). In conclusion, 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

Manuscript received on June 30, 2006.

This work is supported by the European Community, under the Sixth Framework Programme, Information Society Technology, within the IP project "Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery", 2006-2010 [IST-2004-026996 ACGT].

M. Tsiknakis, G. Potamias, K. Marias and S. Sfakianakis are with the Biomedical Informatics Laboratory, Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Crete, Greece (phone: +30-2810-391690; fax: +30-2810-391428; e-mail: {tsiknaki, potamia, kmaria, ssfak}@ics.forth.gr).

D. Kafetzopoulos is with the Post-genomic Technologies Lab of the Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology - Hellas, Heraklion, Crete, Greece (phone: +30-2810-391594; fax: +30-2810 391101; e-mail: kafetzo@imbb.forth.gr).

A. Analyti is with the Information Systems Laboratory, Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Crete, Greece (phone: :+30-2810-391632; fax: +30-2810-391638; e-mail: analyti@ics.forth.gr).

tools for comparing results and experiments, and a uniform platform supporting the seamless integration and analysis of disease-related data at all levels.

II. A European Biomedical Grid Infrastructure for Clinical Trials on Cancer: The ACGT Vision

Within such a context, the implementation of the EU funded Integrated Project named "*Advancing Clinico-Genomic Trials on Cancer: Open Grid Services for Improving Medical Knowledge Discovery*", with the acronym *ACGT*, has begun.

The ultimate objective of the *ACGT* project is the provision of a unified technological infrastructure which will facilitate the seamless and secure access to multilevel biomedical data, its semantic integration, the discovery and orchestration of advanced tools for analysis and visualization and knowledge discovery and their orchestration in seamless scientific workflows (see Fig. 1).

In so doing, *ACGT* aims to contribute to (a) the advancement of cancer research for revealing the influence of genetic variation in oncogenesis, (b) the promotion of molecular classification of cancer and the development of individualised therapies, and (c) the development of realistic and reliable in-silico tumour growth and therapy response models (for the avoidance of expensive and often dangerous examinations and trials on patients) [3].

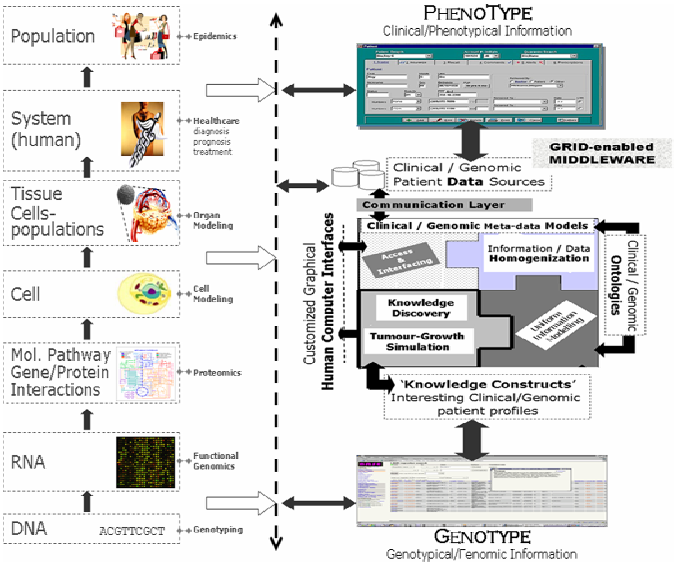

Figure 1: The envisioned ACGT GRID-enabled infrastructure and integrated environment – integration to be achieved at all levels, from the molecular to system and to the population.

The real and specific problem that underlies the ACGT concept is "co-ordinated resource sharing and problem solving in dynamic, multi-institutional, Pan-European virtual organisations". This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or organisations defined by such sharing form what we call the *ACGT* virtual organisation (VO).

In achieving the above objectives, we envisage a need for:

- highly flexible and dynamic sharing relationships. The dynamic nature of sharing relationships means that we require mechanisms for discovering and characterising the nature of the relationships that exist at a particular point in time. For example, a new participant joining a VO must be able to determine what resources it is able to access, the "quality" of these resources, and the policies that govern access;
- sophisticated and precise levels of control over how shared resources are used, including fine-grained and multi-stakeholder access control, delegation, and application of local and global policies;
- sharing of varied resources, ranging from programs and data to computers;
- diverse usage models, ranging from single user to multi-user and from performance sensitive to cost-sensitive.

Consequently, *ACGT* 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 centres 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. Furthermore, *ACGT* intends to build upon the results of several biomedical Grid projects and initiatives, such as the caBIG [4], BIRN [5], MEDIGRID [6], MyGRID [7] and DiscoverySpace [8].

III. The *ACGT* Clinical Trials

It is widely recognised that the key to individualizing treatment for cancer lies in translational research, i.e. in finding ways to quickly "translate" the discoveries about human genetics made by laboratory scientists in recent years into tools that physicians can use to help make decisions about the way they treat patients [9].

The new scenarios of genomic medicine introduce significant new challenges that cannot be addressed with our current methodologies. ACGT focuses on the support of multi-centric, post-genomic translational clinical trials, by creating 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.

Three main clinico-genomic trials (C-GT) have been selected by the project, with the dual purpose of (a) collecting and analysing requirements and (b) performing systems and tools' validation.

1. The first trial – the TOP trial - focuses on ***breast cancer*** (BC) and addresses the *predictive* value of

gene-expression profiling (based on microarrays and genotyping technology) in classifying (according to induced 'good' and 'bad' prognostic molecular signatures) and treating breast cancer (BC) patients.

2. The second trial focuses on ***paediatric nephroblastoma*** or, ***Wilms tumour*** (PN) and addresses the treatment of PN patients according to well-defined risk groups in order to achieve the highest possible cure rates, to decrease the frequency and intensity of acute and late toxicity and to minimize the cost of therapy. The main objective of this trial is to explore and offer a molecular extension dimension to PN treatment harmonized with traditional clinico-histological approaches.
3. The third trial focuses on the development and evaluation of ***in silico*** tumour *growth* and tumour/normal tissue *response simulation* models – in silico tumour growth and simulation modelling (IS-TGSM). The aim of this trial is to develop an 'oncosimulator' and evaluate the reliability of *in-silico* modelling as a tool for assessing alternative cancer treatment strategies.

A. Requirements Engineering – A Scenario Based Approach

The complexity of the domain addressed by the project necessitates that a spiral process of requirements analysis, elicitation, documentation and validation is adopted. Specific techniques, i.e. scenarios and prototyping, elicitation, negotiation and agreement of requirements as well as their validation [10].

On the systems' level, the scenarios guide the specification, the development and the evaluation of the GRID-enabled *ACGT* integrated environment and platform. On the clinical and genomics levels, these scenarios offer clear-cut references for assessing the reliability of the *ACGT*-based technology platform.

A variety of scenarios have been developed by the *ACGT* user community as well as several "technology-driven" scenarios, with the purpose of eliciting requirements and guiding specifications. Such a scenario is presented below. The scenario presents the needs of a researcher testing a hypothesis to explain behavior of non-responders patients who were withdrawn from a given clinico-genomic trial.

In order for this to be achievable the user needs to be supported by the platform in executing the following steps, which constitute the "scenario":

- Identify the TOP trial patients' cases with inflammatory breast cancer that show less than 50% tumour regression and chromosomal amplification in region 11q, who received less than 1 Epirubicine cycle due to serious adverse event allergy in the clinical trial databases of all cancer centers participating in clinical trial.
- Exclude those who show polymorphisms in the specific glucuronidating enzyme of epirubicin UGT2B7
- Query the corresponding genomic databases for the pre-operative and post-operative gene expression data of these patients.
- Normalize the retrieved data, from all participatring in the trial genomic databases, using a selected transformation method.
- Compare with the shown differential gene expression between pre-operative and post-operative data.
- Cluster genes using an appropriate hierarchical clustering method.
- Present the 50 most over-expressed and under-expressed genes.
- Obtain functional annotation for those genes from the GO HUGO and GeneBanks public databases.
- Identify those genes expressed in B-lymphocytes from public GE databases.
- Map those genes into regulatory pathways using a selected visualization tool.
- Finally, get the literature related to kinases present in pathway A and Pathway B and identify their regulatory factors.

IV. THE ACGT ARCHITECTURAL REQUIREMENTS

In responding to these requirements the project focuses on the semantically rich problems of dynamic resource discovery, workflow specification, and privacy preserving distributed data mining, as well as metadata and provenance management, change notification, and personalization. The research and development work in *ACGT* contains the following main components:

- **BIOMEDICAL TECHNOLOGY GRID LAYER:** This layer comprises the basic "*Grid engine*" for the scheduling and brokering of resources. This layer enables the creation of "*Virtual Organisations (VO)*" by integrating users from different and heterogeneous organisations. Access rights, security (encryption), trust buildings are issues to be addressed and solved on this layer, based on system architectural and security analysis.
- **DISTRIBUTED DATA ACCESS:** Provide seamless and interoperable data access services to the distributed data sources, including public databases and in house Clinical Trial Management databases.
- **DATA MINING AND KNOWLEDGE DISCOVERY TOOLS:** The "Data mining and Knowledge Discovery Services" layer includes open data mining and data analysis services. *ACGT* will devote significant effort towards the design, development and deployment of *open*, *interoperable* data mining and analysis software tools and services. The ultimate goal is to offer a GRID-enabled *Knowledge Discovery Suite* [9] for supporting discovery operations from combined clinico-genomic biomedical data.
- **ONTOLOGIES AND SEMANTIC MEDIATION TOOLS:**

Formalised knowledge representations (ontologies) play a central role in the *ACGT* architecture. By building on the various ontologies and controlled vocabularies that have grown over the years for providing a shared language for the communication of biomedical information (e.g., the Gene Ontology (GO), the MGED Ontology, the NCI Thesaurus and Metathesaurus, the UMLS Metathesaurus, etc.), *ACGT* is devoting significant R&D effort to the task of constructing a shared ontology for the disease under investigation.

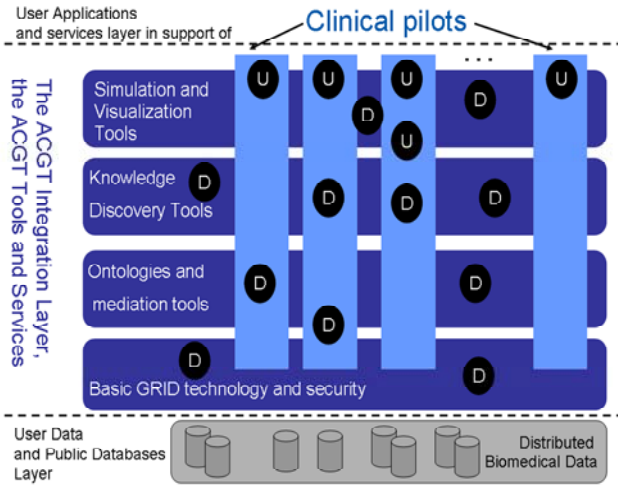

Figure 2: The ACGT architectural layers

- **TECHNOLOGIES AND TOOLS FOR IN-SILICO ONCOLOGY:** *ACGT* will demonstrate its added value for the in-silico modelling of tumor growth and therapy response. The aim been to develop open tools and services for the four dimensional, patient specific modelling and simulation of the biological activity of malignant tumors and normal tissues in order to optimize the spatiotemporal planning of various therapeutic schemes. Ultimately, the aim of this activity is to contribute to the effective treatment of cancer and to contribute to the understanding of the disease at the *molecular, cellular, and higher level(s) of complexity*.
- **THE INTEGRATED *ACGT* ENVIRONMENT:** Integration of applications and services will require substantial meta-information on algorithms and input/output formats if tools are supposed to interoperate. Assembly of tools for virtual screening into complex workflows will only be possible if data formats are compatible and semantic relationship between objects shared or transferred in workflows are clear.

V. R&D CHALLENGES AND THE *ACGT* APPROACH

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, and those to be selected during the lifecycle of the project, Clinical Trials can be demonstrated and evaluated against user requirements defined at the onset of the project.

A critical feature of *ACGT* is to enable semantic interoperability between available data and analytical resources [11]. The key semantic integration architectural objectives in *ACGT* include:

- the development of semantic middleware technology, enabling large-scale (semantic, structural, and syntactic) interoperation among biomedical resources and services on an as-needed basis;
- the development of a shared mediator ontology, the *ACGT* Master Ontology, through semantic modeling of biomedical concepts using existing ontologies and ontologies developed for the needs of the project;
- the mapping of local conceptual models (clinical, genomic) to the shared ontology while checking consistency and integrity of the mapped information;
- the development of a semantic-based data service registry to allow advertisement and discovery of data services on the grid. Such a registry will allow *ACGT* clients to discover data services that have a particular capability or manage a particular data source;
- the semantic annotation and advertisement of biomedical resources, to allow metadata-based discovery and query of biomedical resources by users, tools, and services;
- the descriptions of wet lab experiments, in silico experiments, and clinical trials augmented with metadata so as to provide adequate provenance information for future re-use, comparison, and integration of results.

Some of the challenges facing *ACGT* and the approach taken in tackling those challenges are briefly described in the following sections.

A. *The ACGT Grid Layer*

We have selected the Globus Toolkit for the implementation of the grid middleware for building our open grid layer. The Globus Toolkit [12] is an open source software toolkit developed by the Globus Alliance and many others. The Globus Toolkit provides grid services that meet the requirements of the Open Grid Service Architecture and are implemented on top of the Web Service Resource Framework. It includes software for security, information infrastructure, resource management, data management, communication, fault detection, and portability.

The most important components of the Globus Toolkit involved in our envisaged grid system is WS-GRAM (Web Services – Grid Resource Allocation & Management) for job execution, MDS4 (Monitoring & Discovery System) and GSI (Grid Security Infrastructure). Other technologies that will be included in *ACGT* are Globus security and OGSA-DAI as a grid data layer for exposing data services. The

OGSA-DAI data service is responsible for accessing and retrieving clinical and genomic information from the corresponding information systems [13].

B. Semantic Data Integration and the ACGT Master Ontology

In recent years, there has been an enormous growth in the number of publicly accessible databases on the Internet. All indications suggest that this growth will continue in the years to come. Semantically coherent and integrated access to these data presents several complications and problems [14].

The first complication is *distribution*. Many queries will not be answered by providing data from a single database. Useful relations and data may be broken into fragments that are distributed among distinct databases. Database researchers distinguish among two types of fragmentation; horizontal and vertical fragmentation. Distributed databases can exhibit mixtures of these types of fragmentation. Later, we will see more information about these types of fragmentation and will discuss more about the problem that this kind of division raises.

A second complication in database integration is *heterogeneity*. This heterogeneity may be **notational** or **conceptual**. *Notational heterogeneity* concerns the access language and protocols. One source might use a DBMS using a concrete query language while another source uses the same DBMS but with a different query language. A third example might use, too, a complete different DBMS and query language. This sort of heterogeneity can usually be handled through commercial products.

However, even if we agree that all the databases in a distributed system use a standard hardware and software platform, language and protocol, there can still be a *conceptual heterogeneity* as differences in their relational schemas and vocabulary. Distinct databases may use different words to refer to the same concept, and/or they may use the same word to refer to different concepts. Reassembling the distributed fragments of a database in the face of heterogeneity might prove difficult.

The process of heterogeneous database integration may be defined as "the creation of a single, uniform query interface to data that are collected and stored in multiple, heterogeneous databases."

Classical approaches to database integration [14] include techniques such as wrappers or virtual conceptual schemas. Ontologies are a relevant method for database integration and, in fact, many current projects and proposals are evolving towards ontology-based methods [15]. By using these ontology-based approaches, developers can map, for instance, objects belonging to a specific database to concepts of a shared ontology or biomedical vocabulary.

There are numerous definitions for the term "Ontology". One of the most cited is the one given by Gruber: "An ontology is an explicit specification of a conceptualization" [16]. An ontology can also be described as what it provides: a conceptual framework for a structured representation of the meaning, through a common vocabulary, of a given domain (e.g. medical ontologies describe certain medical domain), specifying concepts, relationships between such concepts and axioms in a formal manner.

Our approach to heterogeneous data integration is based on a mediator-wrapper architecture enabled by the use of ontologies/metadata. In particular, the mediator will integrate heterogeneous data sources (which in the context of *ACGT* are clinical and genomic databases, public databases, web sources, web data services) by providing a virtual view to their data. Users (including *ACGT* tools or services) forming queries to the mediated system do not have to know about data source location, schemas, or access methods, since the system presents one shared mediator ontology (the *ACGT* Master Ontology on Cancer) to the users, who are forming their queries using its terms.

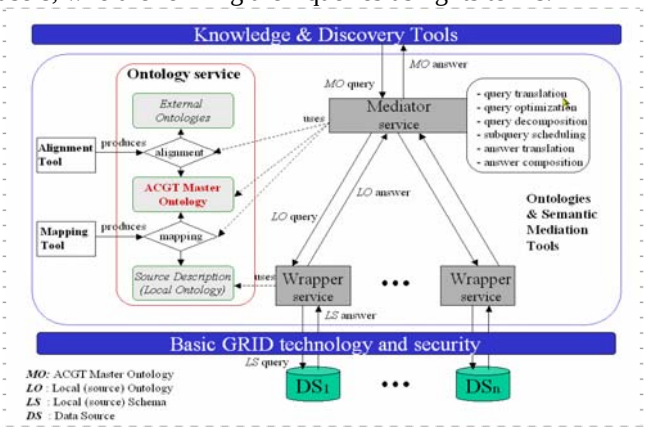

Figure 3: Approach for the heterogeneous, multi-level biomedical data integration in ACGT

In order for the mediator to integrate the various heterogeneous data sources, their object models, terminologies, embedded domain ontologies, hidden semantic information, query capabilities, and security information are analysed. Based on this analysis, a source description is been defined consisting of a local ontology along with a set of metadata, specifying query capabilities and security information.

C. Semantic Discovery of services

A critical requirement of the ACGT infrastructure is that it supports the ability of researchers to discover available resources. The ACGT architecture enables this ability by taking advantage of the rich structural and semantic descriptions of data models and services that are available. The overall architectural considerations for service advertisement and discovery are shortly discussed.

Each service is required to describe itself using a standard service metadata. When a Grid service is connected to the ACGT Grid, it registers its availability and service metadata with a central indexing registry service (the Globus Toolkit's Index Service). This service can be thought of as the "yellow pages" and "white pages" of ACGT. A

researcher can then discover services of interest by looking them up in this registry using high-level APIs and user applications [17].

ACGT employs standards for service metadata to which all services must adhere. The basic metadata supported is the Common Service Metadata standard that every service in ACGT is required to provide. This metadata contains information about the service-providing cancer center, such as the point of contact and the institution's name. Extending beyond this generic metadata there are two standards that are specialized depending on whether a data or analytical service is described. The Data Service Metadata details the domain model from which the Objects being exposed by the service are drawn. Additionally, the definitions of the Objects themselves are described in terms of their underlying concepts, attributes, and associations to other Objects being exposed.

Similarly, the Analytical Service Metadata details the Objects using the same format as the Data Service Metadata. In addition to detailing the Objects definitions, the Analytical Service Metadata defines the operations the service provides. The input parameters and output of the operations are defined by referencing the appropriate Object definition. In this way, both the data and analytical services fully define the domain objects they expose by referencing the relevant concept in the ACGT Master Ontology.

The discovery API and tools of ACGT allow researchers to query the Index Service for services satisfying a query over the service metadata. That is, researchers can lookup services in the registry using any of the information used to describe the services. For instance, all services from a given cancer centre can be located, data services exposing a certain domain model or objects based on a given semantic concept can be discovered, as can analytical services that provide operations that take a given concept as input.

D. E-Science Workflows

The Workflow Management Coalition [18] defines *workflow* as "The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules".

At the current pace of information production in biomedicine there is an unprecedented demand for extraction and processing of knowledge. This is more than evident in various scientific fields such as molecular biology, high energy physics, and astronomy. Consequently, scientific workflows have been proposed as a mechanism for coordinating processes, tools, and people for scientific problem solving purposes [19]. They aim to support "coarse-granularity, long-lived, complex, heterogeneous, scientific computations".

With the advent of Grid technologies the need for the development of workflows technologies that take advantage of the GRID infrastructure and resources has emerged. A Grid workflow can be defined as an automation of a Grid process, in whole or part, during which documents, information or data are passed from one Grid service to another for action, according to a set of procedural rules.

In providing an open, integrated environment for Clinical Trial management using workflows, *ACGT* focuses on the integration of a vast range of resources in terms of data and applications. These resources may be within an organisation, for example in-house systems at a given clinical research organisation or local tools developed within an academic research group, or may be external services delivered by a public body or accessed across an extranet.

The *ACGT* project has identified key user needs wrt to clinical trial workflows. These are:

- *Workflow lifecycle:* Use of a workflow as part of a scientific endeavor requires support for the workflow lifecycle, i.e the construction, enactment, monitoring, evaluation, and persistence of workflows.
- *Semantic description of workflows:* The workflows (and resources) for a particular clinical trial will not necessarily be known a-priori. Specification at a semantic level of the resources and activities required will allow dynamic discovery of suitable resources (in the context of a European open federation of resource providers and resource consumers) and workflows.
- *Workflow provenance:* Use of workflows as part of scientific activity often require provenance data [20] to be kept about activities performed during workflow execution (e.g. details of specific service providers, versions of data and tools involved, etc).

The *ACGT* master ontology, along with additional service/workflow metadata and ontologies, is used for annotating services and ready made workflows (involved in wet lab experiments). Service and workflow annotations provide information regarding the service interface, functionality, provider, quality of service, etc. Annotated services and workflows are registered in the service/workflow registry, organized in classes. Based on these annotations, and assisted by the service and workflow discovery module, the user should be able to semi-automatically compose new scientific workflows.

The use of ontologies and metadata is graphically shown in Fig. 4.

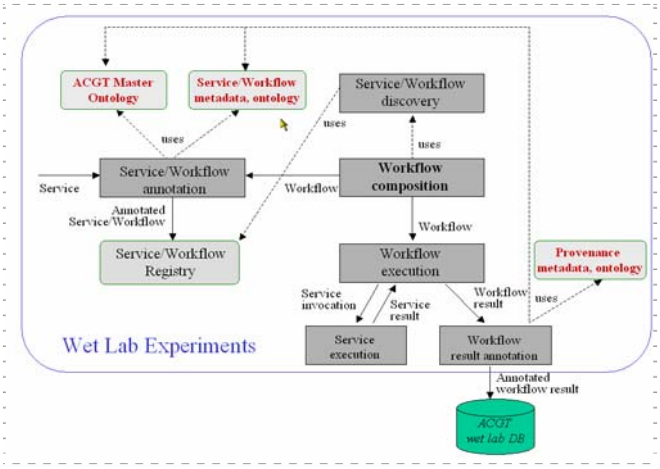

Figure 4: The use of ontologies and metadata in wet lab experiments

VI. Discussion and Conclusions

ACGT brings together internationally recognised leaders in their respective fields, with the aim to deliver to the cancer research community an integrated Clinico-Genomic ICT environment enabled by a powerful GRID infrastructure.

In achieving this objective *ACGT* is progressing with the implementation of a coherent, integrated workplan for the design, development, integration and validation of all technologically challenging areas of work. Namely: (a) **GRID**: delivery of a European Biomedical GRID infrastructure offering seamless mediation services for sharing data and data-processing methods and tools, and advanced security; (b) **Integration**: semantic, ontology based integration of clinical and genomic/proteomic data - taking into account standard clinical and genomic ontologies and metadata; (c) **Knowledge Discovery**: Delivery of data-mining GRID services in order to support and improve complex knowledge discovery processes, (d) **e-science Workflows**: a workflow environment and tools for the visual, semantics-based discovery of resources and their seamless orchestration into complex e-science workflows, for their subsequent execution.

The technological platform is to be validated in a 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.

Finally, it is worth mentioning that *ACGT* promotes the principle of open source and open access, thus enabling the gradual creation of a European Biomedical Grid on Cancer.

Acknowledgment

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

References

[1] Editorial. Making data dreams come true. (2004), Nature 428, 239.
[2] I. Foster. The Grid: A New Infrastructure for 21st Century Science. (2002), *Physics Today*, 55(2):42-47.
[3] C. Sander. Genomic Medicine and the Future of Health Care. (2000), Science, 287(5460): 1977-1978.
[4] D. Fenstermacher, C. Street1, T. McSherry, V.l Nayak, C. Overby, M. Feldman. The Cancer Biomedical Informatics Grid (caBIG™). (2005). Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China.
[5] http://www.nbirn.net/
[6] M. Bertero, P. Bonetto, L. Carracciuolo, L. D'Amore, A. Formiconi, M. R. Guarracino, G. Laccetti, A. Murli, G. Oliva. "MedIGrid: A Medical Imaging Application for Computational Grids," (2003), International Parallel and Distributed Processing Symposium; pp. 252.
[7] http://www.mygrid.org.uk
[8] http://www.bcgsc.ca/discoveryspace/
[9] L. Hood, et al., "Systems biology and new technologies enable predictive and preventative medicine". (2004), Science 306: 640-643
[10] A. Sutcliffe, "Scenario-Based Requirements Engineering". In 11th IEEE International Requirements Engineering Conference (RE'03), 2003, p.320.
[11] G. Kickinger, P. Brezany, A Min Tjoa, J. Hofer, "Grid Knowledge Discovery Processes and an Architecture for Their Composition" (2004). In: IASTED Conference 2004, Innsbruck, Austria, February 17-19, 2004
[12] The Globus Alliance, www.globus.org
[13] The OGSA –DAI project, www.ogsadai.org.uk
[14] W. Sujanski, Heterogeneous Database Integration in Biomedicine. (2001), Journal of Biomedical Informatics; 34 (4):285-298.
[15] J. Köhler, S. Philippi, M. Lange. SEMEDA: ontology based semantic integration of biological databases, (2003), Bioinformatics 19(18):2420-2427.
[16] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications", (1993), Knowledge Acquisition, 5(2), 199-220.
[17] A.C. Von Eschenbach, A.C. "A vision for the National Cancer Program in the United States", (2004), Nature Rev. Cancer, vol.4, pp.820-828.
[18] Workflow Management Coalition, http://www.wfmc.org/
[19] P. Munindar Singh, A. Mladen, "Scientific Workflows: Scientific Computing Meets Transactional Workflows" (1996), NSF Workshop on Workflow and Process Automation in Information Systems: State-of-the-art and Future Directions, May 1996. http://www.csc.ncsu.edu/faculty/mpsingh/papers/databases/workflows/sciworkflows.html
[20] M. Greenwood, C. Goble, R. Stevens, J. Zhao, M. Addis, D. Marvin, L. Moreau, T. Oinn. Provenance of e-Science Experiments - experience from Bioinformatics, (2003), Proceedings UK OST e-Science 2nd All Hands Meeting 2003, Nottingham.