The ACGT Exploitation Plan Update 3 – Final (2010)

Project Number: FP6-2005-IST-026996

Deliverable id: D 16.5

Deliverable name: The ACGT Exploitation Plan

Date: 28 July, 2010

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery
Document id:	D 16.5
Document name:	The ACGT Exploitation Plan – Update 3 Final (2010)
Document type (PU, INT, RE)	RE
Version:	Final
Date:	28.07.2010
Authors: Organisation: Email Address:	Ourania Konstanti, Chris Andronis, Andreas Persidis, Biovista andreasp, o.konst, candronis - @biovista.com

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

ABSTRACT: This document is the final update of the ACGT Initial Exploitation plan.

The present version of the exploitation plan focuses on the exploitable results produced by the consortium, and on exploitation activities undertaken and results obtained.

The presentation of these results is done in the context of previously defined exploitation frameworks options so as to provide consistency and continuity with the original plan itself.

Partner specific exploitation is also presented in detail since it is here that significant progress has been made.

To increase the utility of this document to non-ACGT members, the report compares original success criteria, and their metrics with actual results obtained and where appropriate draws recommendations that will be hopefully useful to readers of this document.

Where specific goals have not been met or partially met, lessons learned are presented again in the hope that they can act as guiding principles for future efforts in the general domain of ACGT.

A set of appendices finally presents details of many follow on projects and initiatives mentioned in the main part of the document.

KEYWORD LIST: exploitation plan, report on activities, exploitation results, exploitation lessons learned

MODIFICATION CONTROL			
Version	Date	Status	Author
0.5	10.07.2010	1st Draft	Ourania Konstanti, Chris Andronis, Andreas Persidis
1.0	16.08.2010	Final	See list of contributors

List of Contributors

- Manolis Tsiknakis (FORTH)
- Nikolaus Forgo (University Hannover)
- Norbert Graf (University Saarland)
- Christine Desmedt (Jules Bordet Institute)
- Brecht Claerhout (Custodix)
- Alberto Anguita (UPM)
- Oswaldo Trelles (UMA)
- George Stamatakos, (NTUA ICCS)
- Dimitra Dionisiou (NTUA ICCS)
- Anca Bucur (Philips)

Contents

1 Exploitation Outcomes Executive Summary

1.1 Overview

ACGT aimed to provide open source tools, resources and infrastructure that in principle target all major stakeholders with an interest in clinical trials and research in a post genomic environment. The project covered a very wide scope of activities and needs, often contradicting (e.g. the need for easy access to clinical data and the requirement for protection of personal information), and attempted to address these with technological solutions that in some cases are based on shifting or incomplete standards and immature underlying IT platforms.

It terms of results the following has been achieved:

1. The main concept of the project, namely to support clinical trial research in multi-centric, multi-platform, international environments has been proven to be technically feasible.
2. Much of the underlying infrastructure developed (including the data repository, security and end-user services levels to name but a few) is technically complex and in certain cases (e.g. the Grid) would probably not be chosen again for the original purpose. On the other hand a number of individual modules are already being used in non-ACGT applications with significant success (see individual partner exploitation results).
3. In addition to the technical results, there has been significant work carried out in the legal and ethical fields surrounding clinical trial design and running. While we do not expect these issues to be considered resolved, especially at the EU level, the results produced represent a significant contribution to the community on which future discussions at the policy and indeed research level could be based.
4. As part of the aborted ACGT Competition (see Deliverable D16.4) an extensive corpus of training materials has been produced. This can be used by newcomers who will wish to use ACGT results or pick up and continue ACGT module development from where the project has left them.
5. The ACGT Video, which represents another “training material”, was developed with the supervision of WP16 management. Available in electronic format and also in 500 DVD copies for immediate distribution, the video provides a ‘popular science’ level introduction to the problem area addressed, the solutions explored and developed as well as challenges remaining. It is expected to increase awareness of the ‘ACGT world’ and will be used to create momentum early on for the p-medicine project, which is one of the follow on projects spawned by ACGT.
6. ACGT has indeed spawned a number of follow on projects (some already successful, some in the preparation stage) and equally importantly specific modules developed for the project are already in use in successful commercial applications.

On the “lessons learned” front ACGT has also generated findings, which, at the top level, can be summarised as follows:

1. As already identified during the project, the technical complexity and scope of the solutions and tasks covered create a steep learning curve as well as require a significant commitment of time from any third party who would be interested to use and benefit from ACGT results. This was clearly seen during the preparation of the ACGT Competition (see deliverable D16.4) but also born out by take up in all major end user sites, especially when ACGT is seen as a whole solution. On an individual

module basis however, the situation is much better with successful uptake in at least 2 instances.

2. Related to the above is the requirement by third parties for assurances for the continued existence and support of the ACGT infrastructure after the official end of this project in order for them to commit any personal resources on learning and using ACGT outcomes. Results here were not as hoped for, however the success of certain project applications (such as p-medicine, ENCCA and INTEGRATE) provides a partial answer to the extent that much of the ACGT will survive and be further developed and assuming that it will be offered as a springboard for that work and made available to end users as early as possible.
3. It is felt that the lack of one or more *compelling* end user applications does not help with the uptake of the ACGT platform as a whole, a difficult task in its own right. Creating more than one internal champion who will dedicate their time and effort to use and show concrete benefits of any technological novelty in actual working environments is a major consideration that must be addressed in depth by all future R&D efforts of this type. Moreover this must be done while ensuring direct use by individual modules of the infrastructure being promoted in a triple-win arrangement between the end users, the service/module developers and the infrastructure managers.
4. On both the legal and ethical levels which ACGT attempts to address, one of the conclusions is that on many of the core issues, there is significant fragmentation at the European level which of course presently makes it quite difficult to deploy an international multi-centric infrastructure that meets all national requirements.
5. Exploitation opportunities and challenges were organized according to three distinct viewpoints of ACGT:
    - ACGT as an infrastructure: while effort was dedicated by the consortium to this goal, results
    - ACGT as an integrated environment of resources, tools and services for end users: some success has been achieved in this respect
    - ACGT as a research project: considerable success has been achieved in this respect with a number of follow-on efforts launched, some of which (e.g. p-medicine) have already been approved for financing

In conclusion exploitation of ACGT work has produced a mixed set of results, with some very good outcomes in terms of individual modules and follow-up projects and some less than hoped for outcomes at the level of MIS infrastructures. Given the complexity of the task, the time involved and the host of external non-technical factors that were not the objective of ACGT this might have been expected. Through the lessons learned section of this document, we aim to capture these experiences, making them available to other strategy designers and policy makers who wish to advance this field in the future.

1.2 Purpose of the Document

The purpose of this document is to present the outcomes of the exploitation activities of the partners and the consortium as a whole, report on the results of specific activities and initiatives foreseen in earlier versions of the plan and discuss follow-on options and opportunities that have materialized as a consequence of these activities.

A second important goal of this document is to present to interested parties (be they other research consortia, policy makers or other stakeholders) the lessons learned by the consortium and individual partners in pursuit of their exploitation activities. Specifically this

document aims to cover, where applicable, experiences gained from individual partner activities as well as from the project as a whole in connection with infrastructure based projects in the area of clinical trials support and also more generally, life science related IT systems.

1.3 Who is this document for?

This document summarizes the work, experiences gained and lessons learned by the ACGT consortium in connection with the exploitation of project results. As such it targets policy makers, researchers as well as IT managers who are active in the area of Medical Information Systems and the support of multi-data, multi-centric IT systems that are deployed in a clinical or more generally health related R&D environment.

1.4 Structure of the Document

This document is arranged in 4 main chapters, as follows:

Chapter 2 presents an overview of the exploitation results, organized in terms of the three options that have been considered by the consortium, namely

- ACGT as an IT infrastructure
- ACGT as a set of resources and services for researchers and clinicians, and
- ACGT as a research project in the field of post-genomic clinical trials support plan proposed for the period 2009-2010 taking into account findings and the work carried out in the third year of the project.

It then presents exploitation activities and results in each of these categories.

Chapter 3 focuses on specific partners and reports on their individual achievements and activities in exploiting their work in the project.

Chapter 4 discusses lessons learned from the exploitation efforts of partners and reflects on the risk factors identified in the previous versions of the Exploitation Plan, specifically whether they have materialised or not and again what lessons can be learned from this.

The Appendices provide additional information and details on a number of issues discussed in the main part of the document.

1.5 Document Versions and Update Procedure

This document is the final update to the original exploitation plan published in March 2007 (deliverable D16.1). It presents the exploitation work package results as they stand at the time of publishing.

All communications regarding the current version should be sent directly to Andreas Persidis, Biovista at andreasp@biovista.com and copied to acgt-wp16@inria.fr .

2 Update on Work to Date

When considering the exploitation of ACGT results, the PMB being cognizant of its complex and multifaceted nature, recognized three organizing frameworks that could be used to focus the work carried out. These frameworks regarded:

a. ACGT as an infrastructure

b. ACGT as an integrated environment of resources, tools and services for end users

c. ACGT as a research project

In the sections that follow we present a brief description followed by exploitation activities addressing that framework. Appendix E presents tables with performance criteria and metrics developed in earlier versions of the plan. It extends these tables by introducing appropriate recommendations the consortium feels are appropriate.

2.1 ACGT as an infrastructure

ACGT is in large part an IT infrastructure project. One of its aims has been to create an integrated environment for the running of services and the exchange of data and other resources based on a layer sitting, architecturally, on top of grid technologies. This main aim has indeed been achieved on a technical basis, thus demonstrating that (a) such an infrastructure is technically possible and (b) laying the foundation for follow on work that will further promote the ACGT design principles and ultimately lead to the broad availability and use of ACGT type services and resources by the target end-user communities.

In terms of the exploitation of the ACGT infrastructure the consortium has had mixed results. While the project has achieved wide visibility in the international community and is now one of the major options available to interested parties, uptake has been less than hoped for in the sense that there is no single end user installation that has deployed ACGT in its entirety, nor are the numbers of third party adopters significant. Appendix E repeats the success criteria that had been set in previous versions of the exploitation plan and expands the relevant table by discussing lessons learned and reasons for the degree of achievement or non-achievement of each specific criterion.

An overall conclusion is that the infrastructure as a whole has not been available for that long (little over the last 6-10 months of the project) in a fixed version that offers developers and end-users something relatively stable to work with. Even as this stability was achieved, questions of complexity, availability and maintenance after the end of the consortium were often directly stated or tacitly implied as reasons for its non-adoption.

On the other hand many of the important “peripheral” aspects of ACGT represent significant contributions to the community and have been exploited both in an educational and a commercial context. These include:

- The entire work on the legal and ethical issues surrounding the use and exchange of potentially sensitive personal information for scientific and research purposes
- The work on data security and anonymization

Each is described in the sections that follow.

Exploitation of Legal and Ethical results of ACGT

The ACGT project offers an interesting case study for legal education in data protection and intellectual property issues. This is of benefit for law schools and students (target group).

Additionally, the data protection and intellectual property framework developed can serve as a basis for further projects in the field of medical research on European level. Interested in

the findings would be other research projects as well as the legal/ethical community (target group). Furthermore, developing the data protection and intellectual property framework lead to re-thinking some of the underlying principles. Therefore, the legal research community could benefit from the project (target group).

For the legal work within ACGT it was, in a first step, very important to come to a common understanding of the legal terms. One example for this is the difference between pseudonymous and anonymous data and what this means in the context of privacy. It turned out that the more interdisciplinary and international projects are, the more it is essential to first agree on the legal concepts and ideas behind in order to understand what can be realised and what cannot from a legal and ethical point of view.

Furthermore, the data protection framework developed for ACGT is a concept that can be built on and adapted for other research projects. It involves the concept of a Trusted Third Party as data custodian and in internal data protection authority. As it is based on the European Data Protection Directive, it is adaptable for other European research projects that have to deal with data protection issues as well as for research projects in all European Member States.

The project has shown that it is of crucial importance to involve a legal advisor at any state in the discussion, especially at the very beginning when the data flows are defined, but also later in the project whenever data shall be transferred from one partner to another. The legal advisor is able to assure compliance with the data protection framework, i.e. check whether the data has been pseudonymised or even anonymised, whether all necessary contracts by the parties involved have been signed and whether the informed consent of the patients exists.

Besides these legal requirements it is from an ethical point of view essential to involve the patients. Autonomy of patients has to be guaranteed. It has to be ensured that all patients consented to the use of their data for research purposes. Moreover, the purposes of data processing have to be defined precisely in order to work with valid informed consents of the patients participating.

Exploitation of Data security and anonymization work

A major exploitation result in this context has been the establishment of the Center for Data Protection (CDP). The CDP is a fully operational entity providing services to 3rd parties. Further exploitation of the CDP at a European level will be achieved through a follow-up project (a support action funded under the HEALTH FP7 programme) whose objective is to provide legal, ethical and security related “services” to all EU cancer related funded projects.

Specifically, the CONTRACT project – through the CDP established in ACGT – will produce the following main outcomes:

1.) Facts and figures about the problems and different approaches in European translational research in matters of informed consent of vulnerable patient groups. This will give European policy makers the information and feedback needed to assess possible changes in the fragmented normative regime of informed consent.
2.) An analysis of the legal and ethical reasons for these different approaches and the role played by EU-legislation.
3.) Identification of good practices in data protection related issues of informed consent in European translational projects with vulnerable patient groups with transferability for translational research projects in general
4.) A policy-oriented study and set of recommendations aimed at EU policy makers regarding possible European coordination measures to optimize the handling of

informed consent for care, research and data protection reasons in Europe

5.) A helpdesk offering concrete advice, FAQs and guidelines on issues of the best possible way of procuring informed consent to protect patient rights and technical solutions related to security for supporting translational research in a transnational environment.

It is worth stating that already more that 20 EU funded projects in the cancer domain, and other initiatives – such as ECRIN – have agreed to capitalize on these services.

2.2 ACGT as an integrated environment of resources, tools and services for end users

As defined in our original exploitation plan, in addition to an infrastructure, ACGT aims to offer non-IT expert end users (namely clinicians, bio-researchers and patients) a set of resources and services that support current tasks and processes that are employed in their working environments (such as the design and management of a clinical trial).

The collaboration with EORTC has been a major success since it encompasses ACGT in its entirety. At the level of resources, the ACGT Master Ontology is also being successfully utilised by third parties and of course plays a central role in the Obtima tool.

ENCCA

ENCCA is a Network of Excellence funded during the October of 2009 call for proposal of the HEALTH programme, under the action line "HEALTH.2010.2.4.1-3: Structuring clinical research in paediatric and adolescent oncology in Europe."

ENCCA's objective is to contribute in making the European Paediatric Medicine initiative a success for children with cancer and to move on from the historical situation where standard treatments and recommendations for the commonly used anticancer drugs in children (dose, schedule, safety, toxicity profile, efficacy, pharmacokinetics) were established by the paediatric oncology community while 50% of these drugs remain unlicensed for paediatric use according to the pharmaceutical rules.

Thirty (30) of the most influential and established academic clinical research groups on paediatric cancer are partners in this "structuring effort" with an additional forty-two (42) centers participating as non-funded partners.

ACGT is an important technology provider in ANCCA, represented though the participation of FORTH. Specifically, ACGT will contribute to:

- T 1.4.2 'Data-sharing and Bioinformatics Tools' Working Party
- T 1.5.4 Definition of disease-specific core datasets for pooled analysis
- T 1.2.4 Definition and management of standards and workflows
- T 1.2.5 Design and specification of the data exchange and inter-operability infrastructure
- T 1.2.6 Development and set-up of an infrastructure prototype

In the context of these activities which will evaluate alternative technological approaches (e.g. applicability of IHE profiles for pulling data out of Hospital Systems for analysis) the ACGT methodological approach and the ACGT technologies (i.e. the integrated ACGT platform) will be evaluated in the context of a new prospective trial. Namely, in ENCCA's task "T 2.4.3 Implementation of a prospective WT clinical study in the ACGT system", FORTH will lead the implementation of a prospective WT clinical study in the ACGT system. To implement prospective clinical, imaging and biological data collection on patients treated according to the standard arms of the SIOP Wilms tumour 2001 clinical study using the

ACGT platform. This will permit integration of complex data from the current clinical database with existing and prospectively acquired molecular biology data and imaging studies (DICOM data) to underpin identification and evaluation of biomarkers (linked to WP1.5.2). The ultimate aim is to implement a Web-based integrated data collection tool for studies on Wilms tumour patients as a proof of principle for other prospective data collections and clinical trials in patient groups with a very favourable prognosis.

ACGT – EORTC
The main points of the collaboration are as follows:

- A review of the security infrastructure of ACGT by EORTC staff which provided comments from an experienced group of end users (included in deliverable 13.2)
- Provision of CRF from EORTC with the aim of improving the ACGT Master Ontology
- Provision of data from a recently closed EORTC trial which has extensive clinical, imaging and biological data.
- Provision of access to the protocol and CRF of the MINDACT trial so as to see the degree to which ACGT infrastructure meets the needs of this trial.

The ACGT Master Ontology

ACGT project researchers undertook a systematic review of existing domain and upper-level ontologies, as well as of existing ontology design software, implementation methods, and end-user interfaces. This included the study of best practices, design principles and evaluation methods for ontology design, maintenance, implementation, and versioning, as well as for use on the part of domain experts and clinicians. The result of this work has been:

1. the development of a master ontology (the ACGT MO) based on clearly defined principles of ontology development and evaluation;
2. the development of a technical infrastructure (the ACGT Platform) that implements the ACGT MO utilizing independent tools, components and resources that have been developed based on open architectural standards, and which includes an application updating and evolving the ontology efficiently in response to end-user needs; and
3. the development of an ontology-based trial management application (ObTiMA) that integrates the ACGT-MO into the design process of clinical trials in order to guarantee automatic semantic integration without the need to perform a separate mapping process.

2.3 ACGT as a research project

When considering ACGT as an R&D project exploitation is understood to mean follow on or related projects that will continue and/or expand on the work and the areas covered by ACGT. In this respect, there are threfour2.3.1e such outcomes that have been achieved:

- The Gen2Phen project
- The Tumor Project
- The NeoBig project
- The STARC Initiative
- The p-medicine project

2.3.1 ACGT and Gen2Phen

The main goal of Gen2Phen project is to unify human and model organism genetic variation databases towards increasingly holistic views into Genotype-To-Phenotype (G2P) data, and to link this system into other biomedical knowledge sources via genome browser functionality.

In the context of the Gen2Phen FP7-Health project (http://www.gen2phen.org/) a Grid-enabled G2P knowledge discovery scenario implemented, called the GG2P scenario realised as a Web-services Grid-enabled workflow. The GG2P scenario aims to designate multiple SNP profiles - among a huge number of them, that not only associate with a disease but also exhibit a high discrimination power between different phenotypic classes. An SNP is a single base substitution of one nucleotide with another. It is known that a category of diseases is associated to a single SNP or gene (also known as monogenic diseases). In general, a single SNP or gene is not informative because a disease may be caused by completely different modifications of alternative pathways in which each SNP makes only a small contribution. Most of the complex diseases, including cancer, are characterized by groups of genes with a number of susceptible genes interacting with each other.

The ACGT Grid infrastructure and its workflow editing and enactment environment were utilized for the implementation of the methodology that enables the discovery of genotype-to-phenotype associations and predictive models, and supports G2P association studies. The GG2P workflow was executed on an indicative SNP genotyping experiment (from the ArrayExpress repository) that concerns the hybridization of breast cancer and normal/control tissue samples. Using advanced data mining algorithms the GG2P scenario identified about 100 indicative SNPs that exhibit contrasted homozygosity / heterozygosity profiles, indicating a clear Loss-of-Heterozygosity (LOH) profile, and achieve highly discriminant performance figures for the respective phenotypic classes. The most highly ranked SNPs exhibit clear loss of heterozigosity patterns, a common situation in tumorgenesis. Literature searches provide strong evidence about the biological relevance of the findings – the respective SNP's genomic regions are strongly association with characteristic breast cancer phenotypes.

The ACGT workflow realization of GG2P is founded on a set of appropriately devised and customised Web-services. The GG2P workflow was also extended to accommodate special Web-services that annotate the final resul;ts output that link the identified SNPs with related public databanks, e.g., dbSNP, Ensemble genome browser, HGVbaseG2P, DiseaseCard and Pubmed. The whole exemplified GG2P workflow is illustrated in Fig 2.3.1

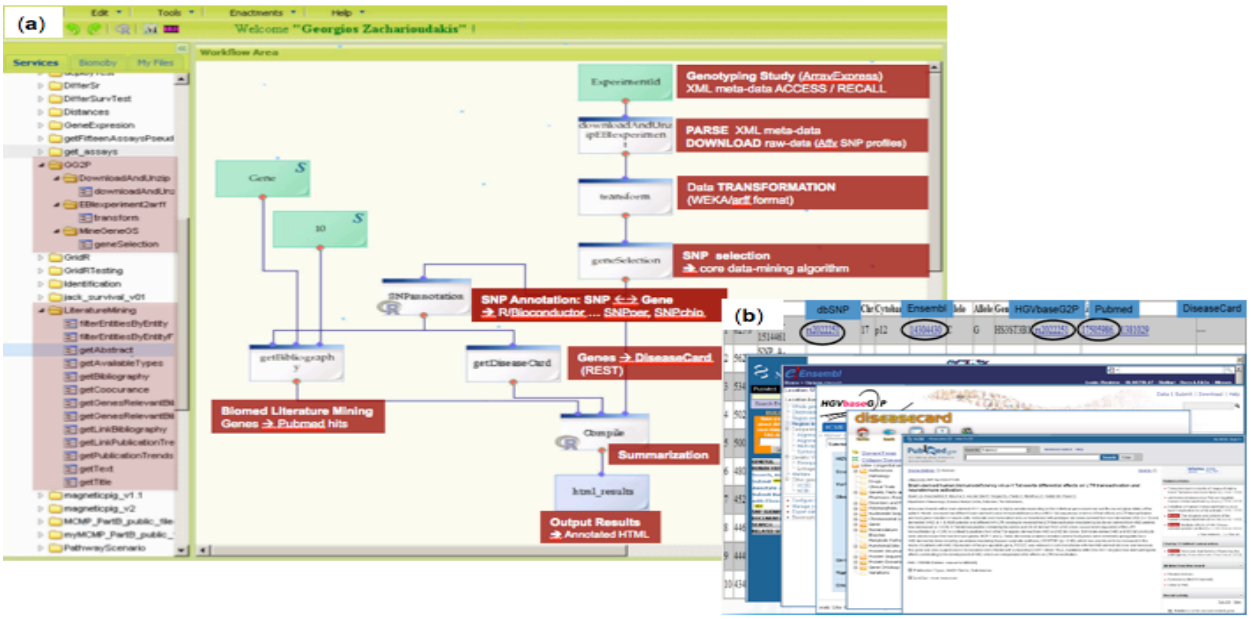


Fig 2.3.1. The GG2P workflow: (**a**) Utilization of ACGT's workflow editing and enactment environment for the discovery of SNP-disease associations (in the context of the GEN2PHEn project); (**b**) Automated linkage of identified SNPs with related public genomic/genetic databanks.

References

1. Koumakis L., Moustakis V., Tsiknakis M., Kanterakis A., Kafetzopoulos D., and Potamias G. (2009). Supporting Genotype–To–Phenotype Association Studies with Grid–Enabled Knowledge Discovery Workflows. Proceedings 31st Annual International IEEE EMBS Conference (EMBC'09), pp. 6958-6962, September 2-6, 2009, Hilton Minneapolis, Minnesota, USA.
2. Koumakis, L., Sfakianakis, S., Moustakis, V., and Potamias, G. (2009). Discovery of Genotype–to–Phenotype Associations: A Grid–enabled Scientific Workflow Setting. *BMIINT: Biomedical Informatics & Intelligent Methods in the Support of Genomic Medicine*, AIAI 2009 workshop, April 24 2009, Thessaloniki, Greece, CEUR Proceedings 475:48-59.

2.3.2 ACGT and the newly funded TUMOR project

The TUMOR project aims at developing a European *clinically oriented* semantic-layered cancer digital model repository from existing EU projects that will be interoperable with the US grid enabled semantic-layered digital model repository platform at CViT.org (Center for the Development of a Virtual Tumor, Massachusetts General Hospital (MGH), Boston, USA) which is NIH/NCI-caGRID compatible.

To achieve these goals, multiscale models/tools developed and data collected within the framework of three ongoing EC funded research projects namely ACGT [Advancing Clinicogenomic Trials on Cancer], ContraCancrum [Clinically Oriented Cancer Multilevel Modeling] and the VPH NoE [Virtual Physiological Human Network of Excellence], in conjunction with models and data from the NIH supported ICBP Program CViT.org will drive the development, optimization and validation of the integrated system.

This interoperable, CViT interfaced, environment will offer a range of services to international cancer modelers, bio-researchers and eventually clinicians aimed at supporting both basic

cancer quantitative research and individualized optimization of cancer treatment. This 'Transatlantic' project will therefore be the starting point for an international validation environment which will support joint applications, verification and validation of the clinical relevance of cancer models.

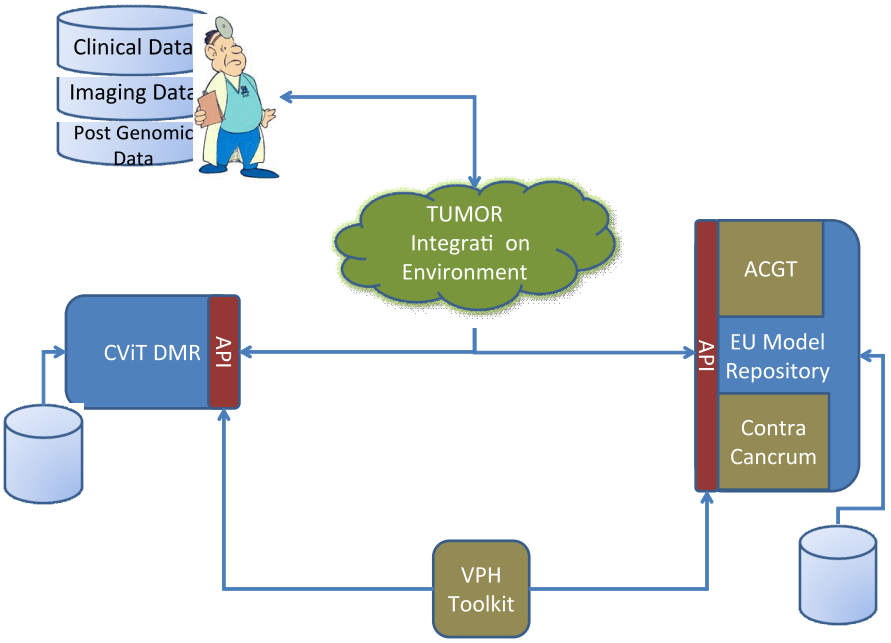


Figure 2.3.2. **High level architecture of the TUMOR environment: Multi-level cancer models/data/tools will be shared within a fully interoperable environment for specific transatlantic scenarios.**

Figure 2.3.2 shows the envisaged architecture of the TUMOR environment. The EU cancer model repository API is integrated with the MGH-CViT DMR API and relevant data sources for deploying specific transatlantic scenarios. VPH model markup language descriptions of employed models will guarantee interoperability. The TUMOR environment will be used to deploy a transatlantic clinical scenario(s) which will be based on combining 'top down' approaches (developed mainly by ICCS), that have been used so far to simulate tumor response to treatment, with 'bottom up' MGH approaches in CViT.

In the relevant EU projects, three different tumor types are currently being analyzed: nephroblastoma, glioma and lung cancer. In ACGT and ContraCancrum the following clinical scenarios are currently being implemented:

Nephroblastoma: individual patients	Simulation of the response to preoperative chemotherapy in
Lung cancer: patients	Simulation of the response to targeted therapies in individual
Glioma:	Simulation of treatment response in individual patients

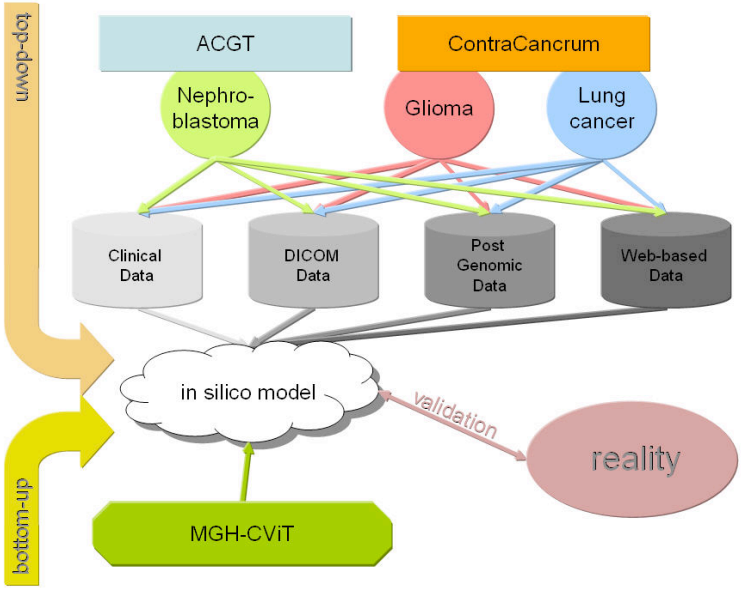


Figure 2.3.3. A general overview of potential scenarios in TUMOR with the use of top-down and bottom-up *in silico* approaches that have been developed by TUMOR participants.

Fig 2.3.3. illustrates the TUMOR Transatlantic, combined scenario (based on glioma data) will be used as a demonstrator case to evaluate cross-repository/platform interoperability and report the clinical results on modeling assisted therapy response prediction.

2.3.3 ACGT and NeoBIG

The NeoBIG program is a research program led by Breast International Group, and aims to accelerate drug and biomarker development in early breast cancer, recognizing that the current drug development process is suboptimal and aims to improve the results of clinical trials. . A durable, multidimensional translational research structure supporting neo-adjuvant trials will be build, sharing strategies, expertise, technologies, methodologies and protocols. In addition this will provide a strong foundation for future adjuvant trials in breast cancer (and research in other cancers). The program should result in a lasting bioinformatics platform for collaboration between cancer research institutes in Europe.

In ACGT we have evaluated whether and how the expertise, and potentially also the tools, developed in ACGT could be used to support a large real-life multi-centric clinical trails programme, such as NeoBIG. Our focus was on the IT needs of such a research programme, specifically with respect to secure privacy-preserving data sharing as these are issues at the core of ACGT. We have tried to answer these questions by first collecting and analyzing the requirements of BIG concerning the data sharing platform needed to support their future clinical trials, and based on that briefly evaluating potential alternatives in which ACGT could support this programme.

During that study, we have concluded that there is a lot of ACGT expertise that could be used for the neoBIG data sharing platform, especially with respect to data storage, management and sharing, and with respect to privacy and security. At the same time, we have understood that while accessing external data out of heterogeneous repositories is highly relevant, there is also very high value in supporting the neoBIG community to build comprehensive datasets including all the wealth of data collected in the neoBIG trials, and to provide infrastructure enabling large scale collaboration and sharing.

In this context, we have defined INTEGRATE, a new collaborative project that aims to build solutions that support a large and multidisciplinary biomedical community ranging from basic, translational and clinical researchers to the pharmaceutical industry to collaborate, share data and knowledge, and build and share predictive models for response to therapies, with the end goal of improving patient outcome.

The INTEGRATE project will develop flexible infrastructure components and tools for data and knowledge sharing and wide scale collaboration in biomedical research. Our infrastructure will bring together heterogeneous multi-scale biomedical data generated through standard and novel technologies within post-genomic clinical trials and seamlessly link to existing research and clinical infrastructures, such as clinical trials management systems, eCRFs, and hospital EHRs, and to relevant external biomedical infrastructures. We will also build repositories of data, annotated models, and metadata and provide tools to extract and manage content, add and update data and models, and link to external sources for complex analyses.

On top of this flexible infrastructure and using the available multi-level data, we will develop and validate models and simulators predicting therapy sensitivity for individual patients. The clinical scenarios will confront the model predictions with the actual outcome of patient therapy.

Next to bringing together data and knowledge, our solutions aim to join a wide multi-disciplinary community of biomedical and clinical researchers committed to work together, to establish common methodologies and clinical protocols, to collaboratively build predictive models, carry out research and select the most suitable integrative workflows. The infrastructure and tools developed by the INTEGRATE project will support BIG to promote in the clinical community new methodologies and define standards concerning the collection, processing, annotation and sharing of data in clinical research and improve the reproducibility of results of clinical trials.

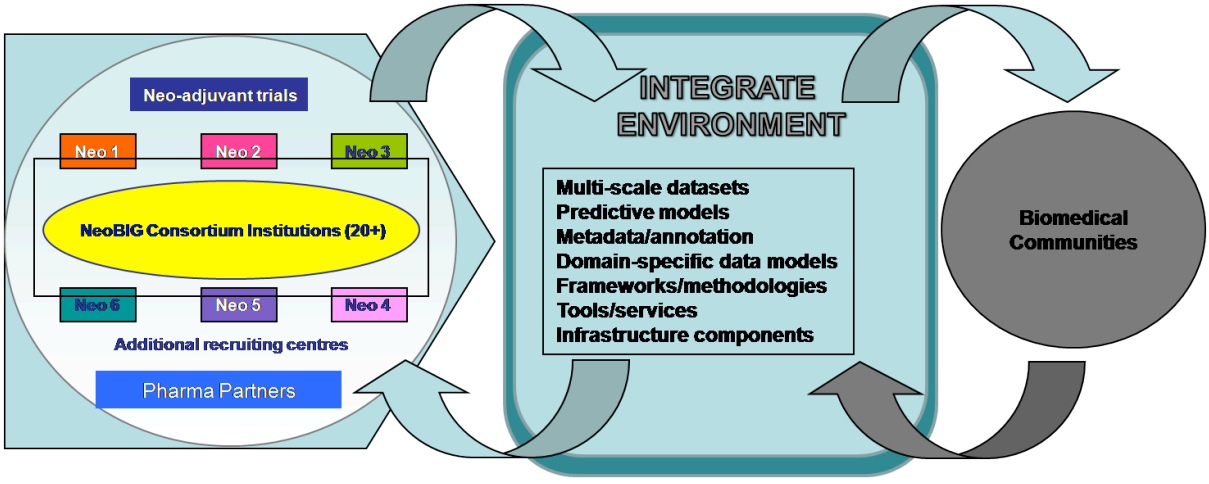


Figure 2.3.4 The INTEGRATE Concept: Sharing and Collaboration among Clinical Research and Biomedical Communities

We will also exploit the unique opportunity of the NeoBIG empowered collaborative environment to combine multi-scale biomarkers (from genetic level to tissue level including imaging biomarkers) in order to define a methodology for improving the prognostic power of currently used practices for assessing neo-adjuvant therapies in Breast Cancer. This will take the form of a 'use-case' VPH scenario emanating from and being deployed within the INTEGRATE environment. The goal is to demonstrate that the predictive power of responsiveness can be enhanced by using multi-scale biomarker signatures. The predictive

models developed, our approach, methodology and the INTEGRATE modelling framework will be contributed to the VPH community.

The development of predictive models in an environment build together with the clinical trials community enables us to benefit of an accelerated adoption process towards the clinical practice. Additionally, by focusing on interoperability with existing clinical infrastructures INTEGRATE aims to reduce the distance between research and clinical practice.

Finally, through the INTEGRATE environment, datasets out of the innovative NeoBIG trials, that as we previously specified will make use of the INTEGRATE sharing and collaboration solutions, will become available in the future to the biomedical community.

2.3.4 The STaRC Initiative:

STaRC is an initiative originally conceived by the University of Saarland (Prof. N. Graf) based on the experience of ACGT and some of the challenges that have been identified by the consortium so far. These challenges refer to the deployment and uptake of ACGT services and infrastructure by third party stakeholders, in particular hospitals and clinicians involved in Clinical Trials. STaRC is intended to be a '**S**tudy, **T**rial **a**nd **R**esearch **C**entre' that will exploit clinically relevant aspects of ACGT. The main tasks of STARC will be:

- Simplification of the clinical trial process
- Patient empowerment
- Combining clinical and molecular biological and genomic data in single patients leading to personalized medicine
- Facilitating translational research
- Continually improving curricula of medical schools and medical education

Appendix A presents in more detail the STaRC initiative.

2.3.5 The p-medicine project

p-medicine (full title: From data sharing and integration via VPH models to personalized medicine) is a proposal submitted by a sub-group of the ACGT consortium (the majority of the ACGT consortium are part of p-medicine). It addresses the strategic objective "ICT-2009.5.3 Virtual Physiological Human – ICT tools, services and specialised infrastructure for the biomedical researcher" and aims to create an innovative computational, service-oriented infrastructure that will facilitate the gradual translation from current medical practices to personalized medicine i.e. from a largely reactive mode of medicine to one that is personalized, predictive, preventive, and participatory.

This change is rooted in *new science*, including the convergence of systems approaches to disease, new measurement, modelling and visualization technologies, and new computational and mathematical tools. While the goal is clear, the path to such advances has been fraught with roadblocks in terms of technical, scientific, and sociological challenges and p-medicine ams to address some of these, building on much of the work carried out in ACGT.

p-medicine has been successful in attracting EC funding which can be taken, at least in part, as a vote of confidence in the work carried out in ACGT.

Fundamentally, p-medicine builds and exploits the majority of the ACGT developments, but the GRID. Instead cloud computing is explored as an alternative to the "tightly coupled" GRID solutions with respect to supporting high performance computing (HPC) requirements.

This is a result of the fact that recently, supercomputing has been made more accessible and affordable through the development of virtualization technology. Virtualization software allows systems to behave like a true physical computer, but with the flexible specification of details such as number of processors, memory and disk size, and operating system. Multiple virtual machines can often be run from a single physical server, providing significant cost savings in server hardware, administration and maintenance.

The use of these on-demand virtual computers is known as cloud computing. The combination of virtual machines and large numbers of affordable CPUs has made it possible for 'mega-scale' computational clusters and advanced, large-scale data-storage systems.

Appendix B presents the project abstract in its entirety.

3 Partner-Specific Exploitation

ACGT is of course a very complex and large undertaking and our main concern is this exploitation as a whole. However, the consortium comprises a number of partners each of which could and does exploit their work in a multitude of ways that can compliment the efforts taken at the consortium level.

With this in mind, we report in this section on specific exploitation activities and results by individual partners. We have selected those results that although primarily benefiting the partner himself, impact ACGT by providing case studies of ACGT compliant resources and services in actual use.

3.1 NTUA-ICCS

NTUA-ICCS exploitation focuses on the further development of the Oncosimulator module. The ACGT Oncosimulator has served as the basis for the evolutionary development of advanced multiscale simulators of tumour response to treatment to be used for the optimization and individualization of cancer treatment through experimentation in silico (i.e. on the computer). To this end various tumour types and treatment combinations have been considered and addressed. Exploitation of the ACGT Oncosimulator has already been made by the EC funded “ContraCancrum” and the EU-US “TUMOR” research and development projects as already mentioned in section 2.3. The Oncosimulator target users include biomedical researchers, medical doctors, students of medicine, biology and (bio)engineering, interested patients, and the general public.

An initial clinical adaptation optimization and partial validation has already been achieved in collaboration with leading clinical centres and medical doctors and biologists. This has set the standards for a systematic validation of future oncosimulators and the subsequent observational substantiation of the emergent multi-disciplinary field of "in silico oncology" Up to now the practical end products of the related actions include clinically adapted multiscale cancer simulation software (glioblastoma, nephroblastoma and breast cancer versions), numerous publications, a short educational video delineating the principles of the Oncosimulator for the radiotherapy case and an international book.

The ACGT Oncosimulator team has acted as an important seed for the formation of an intercontinental research network encompassing EU, US and Japan and aiming at advancing and clinically translating in silico oncology and the notion of the oncosimulator. Many research, development and clinical institutions participate in the network. ACGT was a sponsor of the strategic ICT event entitled: "1st Transatlantic Workshop on Multiscale Cancer Modeling" that was co-funded by the EC and the US National Cancer Institute - NIH and took place in Brussels (Cherlemagne Building) in 2008

http://ec.europa.eu/information_society/events/ict_bio/2008/ta-cancer-wkshp/index_en.htm

ACGT supported the workshop series entitled: "International Advanced Research Workshop on In Silico Onolcogy and Cancer Investigation" (which this year (2010) is an EC sponsored, IEEE-EMBS technically co-sponsored and IFMBE endorsed event).

It should be noted that the Oncosimulator development has a strong basic science component including the formation, formulation and advancement of new quantitative science (with technology playing an important role). A very strict and lengthy clinical validation procedure is necessary before its eventual clinical translation. This leads to a reasonable delay in the eventual clinical translation of the Oncosimulator.

Apart from open access to related simulation software end products (e.g. simulation exploration portals developed in ACGT), the development of the ContraCancrum Oncosimulator, a “second generation” oncosimulator exploiting the expertise accumulated during the ACGT Oncosimulator development, has been integrated into the mid-term planning of the PHILIPS Research company having its base in Hamburg, Germany. PHILIPS Research has been participating in the EC funded ContraCancrum project.

The ACGT Oncosimulator is envisaged to serve as the basis of the evolutionary development of advanced multiscale simulators of tumor response to treatment addressing various tumor types and treatment combinations. This utilization of the ACGT Oncosimulator has already been made clear in the ContraCancrum and TUMOR projects co-funded by the European Commission.

The Oncosimulator target users include researchers, medical doctors, medicine, biology and bioengineering students, interested patients, and the general public.
Within the ACGT framework an initial clinical adaptation optimization and partial validation has been achieved. This has set the initial standards for a systematic validation of oncosimulators and the subsequent observational substantiation of the emergent multi-disciplinary field of in silico oncology.

The ACGT consortium has served as one of the most important seeds for the formation of an intercontinental research "alliance" aiming at advancing both the notions of the oncosimulator and in silico oncology.

Among other practical end products are: simulation software (nephroblastoma, breast cancer versions), the OncoRecipeSheet exploratory software, numerous publications, and a short educational video delineating the principles of the Oncosimulator (radiotherapy case).

3.2 Fraunhofer-Institut Intelligente Analyse- und Informationssysteme IAIS

During the lifetime of ACGT, Fraunhofer IAIS has generated significant know-how in the setup and operation of parallel distributed data mining toolkits in the form of GridR.

Fraunhofer IAIS is exploiting this know-how in its commercial projects, where the GridR technologies prove to be crucial to enable the rapid prototyping and operation of large-scale data mining applications. The following shortly describes a prototypical example of such an application.

GridR was used for the parallelization of a GPS-trajectory mining application that calculates the reach and gross contacts of outdoor poster advertisements (Wegener/etal/2008). The application provides the foundation for price calculations for all outdoor advertisement throughout Germany. In Germany, the outdoor advertisement industry records a yearly turnover of more than 800 million euros. The Arbeitsgemeinschaft Media-Analyse e.V. (ag.ma) – a joint industry committee of around 250 principal companies of the advertising and media industry in Germany – authorized the AGMA project, which provides the foundation for price calculations in outdoor advertisement throughout Germany. In 2006/07 the ag.ma commissioned a nationwide survey about mobile behavior and appointed Fraunhofer IAIS to calculate the reach and gross contacts of poster networks. Using GridR, it was found that a scenario of a simulation of 12 cities could be computed in a few days compared to the hypothetical sequential execution time of roughly one year.

References:

1. Wegener, Dennis and Hecker, Dirk and Körner, Christine and May, Michael and Mock, Michael. Parallelization of R-programs with GridR in a GPS-trajectory mining

application. In Proc. of the ECML/PKDD 2008 First Ubiquitous Knowledge Discovery Workshop (UKD08), Antwerp, Belgium, September, 2008.

3.3 UMA

UMA has been pursuing an active exploitation strategy for most of its research results. The following describes specific exploitation activities related to software developed and used within the ACGT platform:

Metadata repository

Many bioinformatics tools are available on-line but due to rapid developments of new and improved tools, it is necessary to be able to publish and maintain metadata about such tools in a public registry. This metadata can be used to provide several necessary functionalities, for example discovery, invocation and documentation of tools, data persistence systems (data provenance).

Tools available over the internet are particularly useful in biomedical informatics, where large computational resources, both hardware and services; and access to large and constantly updating data collections are basic requirements for solving many typical problems. This is the main motivation for delegation of tool invocation (execution) to remote servers with required computational resources. Common approaches to such distributed tools are web-services (software systems designed to support interoperable machine-to-machine interaction over a network) and workflows (pre-defined organized invocations of web-services).

Here we offer the implementation of a repository for knowledge-discovery-related tool metadata. We use existing formats for tool descriptions and we have also defined a core set of metadata for tools. This metadata is connected to the tool description and improves the possibilities of finding the right tool description among a large set of tool descriptions. The repository allows tool requestors to filter available tools depending on the metadata that the tools are annotated with.

More Info and documentation: http://www.bitlab-es.com/repository/

Modular API

The popularization of the Internet has promoted the development and deployment of a myriad of publicly available web-based resources in a broad range of application domains. Unfortunately this diversification is also present in the access protocols and data formats used, thus creating a barrier for the integrated use of such resources

We have developed a new framework for the development of applications with the main characteristic of be modular and adaptable to different scenarios.

The natural flexibility of the framework allows the connection to different data sources and combines their contents for showing all to the user with a easy to use and powerful interface.

More Info and documentation: http://www.bitlab-es.com/mapi/

Clients jORCA: Easily integrating bioinformatics web services

jORCA is a desktop client able to efficiently integrate different type of web-services repositories mapping metadata over a general definition to support scalable service discovery and to achieve flexible inter-communication between tools.

jORCA manages repositories heterogeneity supported by the Modular-API that provides a

uniform view of metadata (e.g. GRID-based, WSDL–services, BioMoby and others), making the integration of bioinformatics Web-Services easier. This software is intended to facilitate the use of Web-Services offering a friendly tool able to execute and integrate web-services that use different protocols.

More Info and documentation: http://www.bitlab-es.com/jorca/

Magallanes: Multi□architecture resources discovering.

Magallanes is a versatile and platform-independent Java library of algorithms to built- up search engines to help in the discovery of services and datatypes specially oriented to deal with repositories of web-services and associated datatypes. A service or data- type discovery process aims to identify the set of services or data-types that satisfy a given number of constraints from the pool of all the available (e.g. what services are available to process my sequence?). Typically, clients that exploit these repositories supply some mechanism to help in the discovery process, most of them with syntactic orientation such as search for web services by name and also by function or input or output parameters. Text search functionality over the service's name is also frequent in clients. However, for the final user this type of syntactic discovering process is not efficient since it assume a user-knowledge on the names of objects or services. Magallanes offers the necessary architecture for flexible and expandable identification of Web-services in such a way they can increase the client capabilities and enforce the integration process.

The modular API allows connecting Magallanes with external clients in such a way that results can be used to discover and invoke web-services, recover a data-type description, etc. It is also possible directly built-up a desktop application using the GUI functionality. One of the main advantages is the easy way to adapt the client to dealt with repository definition modifications (only parsers need to be adjusted), and its ability to work over new repositories (standardised by the modular API). At present Magallanes is able to access MOWServ, MOBY-Canada, and ACGT repositories. ACGT official Web portal uses Magallanes search engine to provide tool discovering support.

More Info and documentation: http://www.bitlab-es.com/magallanes/

BioDataSF:

A software library for the management, edition and standardization of user data that aims to facilitate data transformation and promote service interoperability by understanding I/O descriptions. The tool, named BioData-SF (Bioinformatics Data Standardization Formatter), is able to use standard data models stored elsewhere to guide the parsing and interpretation of user data (e.g. flat text files or even binary files). The kernel of the system provides a gateway embedding various software for specific datatypes, mixed with an extensible set of rules and regular expressions to manage the automatic identification of data types and perform appropriate transformations. BioData-SF can be used as a software library, or as a plug-in to be embedded in other applications.

More Info and documentation: http://www.bitlab-es.com/biodatasf/

Web Services in ACGT Platform

Data analysis in bioinformatics is typically delegated to external web-services which are designed to support interoperable machine-to-machine interaction over a network.

Two main platforms are common in gene expression analysis: two-channel (competitive) and one-channel analysis hybridization. Two-channel microarrays are typically hybridized with cDNA prepared from two samples to be compared at the same time (e.g. diseased versus healthy samples), while one-channel (typically Affymetrix) contains a set of probes for genes

(technical replications) from which probe-set values are obtained to represent the gene expression. Several analysis methods can be applied to ensure data quality, spot filtering, inter- and intra-slide normalization, replicate resolution, dye-swapping, random error removal and statistical analyses (see Prep+07).

Once data is normalized, it can be used to generate a gene expression matrix. At this stage, more complex data analysis tools can be applied. One example is Engene , which is useful tool for storing, visualizing and processing large sets of expression patterns. In the case of one-channel data, there are several alternatives to normalization of cell data (for example, the Bioconductor R packages Affy (www.bioconductor.org>).

Functionality from Prep, Engene and the quantile normalization implementation has been provided as ACGT and BioMoby compliant web-services and registered in the different metadata repository.

More Info and documentation: http://www.bitlab-es.com/gews/ **List of services:**

3.4 UPM

In addition to the Semantic Mediation modules developed, two important exploitable results developed by UPM are a mapping format that addresses semantic heterogeneities as well as a mapping API

***ACGT Mapping Format* (developed in collaboration with FORTH)**

To support the translation of queries that takes place in the SM, an XML format for describing the correspondences between terms in the ACGT MO and terms in the underlying data sources was created. For each database to be integrated through the SM, an XML file sticking to this format must be created (commonly named “mapping file”). To develop this format, a study of the most common cases of heterogeneities among semantically equivalent databases was carried out. The result is a format that covers the possible semantic heterogeneities between a pair of RDF-based data sources.

ACGT Mapping API

The Java API allows the programmatic creation and exploration of mapping files in the above described mapping format. This API allows other client tools of the SM (such as the Obtima Trial Builder system) to dynamically create and submit mapping files for new clinical trial databases. It is an important resource that will be used in follow on projects and tools (like Obtima).

3.5 BIOVISTA

At the inception of ACGT, it was decided to include Biovista’s literature based mining and discovery technology as one of the services that would be available to end users, namely clinical trials doctors and researchers. The idea was that these end users would benefit from this technology in discovery intensive tasks such as understanding disease mechanisms, and predicting biomarkers or adverse events of drugs. At that point in time, there existed some initial evidence that literature on existing scientific knowledge could indeed be used as a data source for discovery and the creation of new valid hypotheses. Based on this, one of Biovista’s goals within ACGT was to confirm both these claims, namely that (a) literature could indeed be used by predictive algorithms and (b) that such algorithms would be useful to practitioners in a research and clinical trials context. A second goal was to identify other uses of such a capability such as supporting sales efforts by providing personalized or other high value information to potential customers that would help turn them into actual customers. A final goal was to see if and how literature as a resource can be combined with

other resources (such as patient records) to arrive at better therapies and eventually in the more distant future more personalized medicine.

Making these technologies available to third parties in new ways, including via ACGT compatible formats, has allowed all of the above goals to be approached and in what follows we present findings as well as exploitation results achieved.

The Literature mining and discovery platform

Biovista's literature mining and discovery platform is based on the premise that by connecting seemingly disparate chunks of knowledge and information, as these are reported in the scientific literature, it is possible to answer scientific questions and generate new knowledge that is not directly reported in the corpus used; in other words to support scientific discovery.

The technology platform that had been already developed, was based on a number of assumptions and scientific choices as described below:

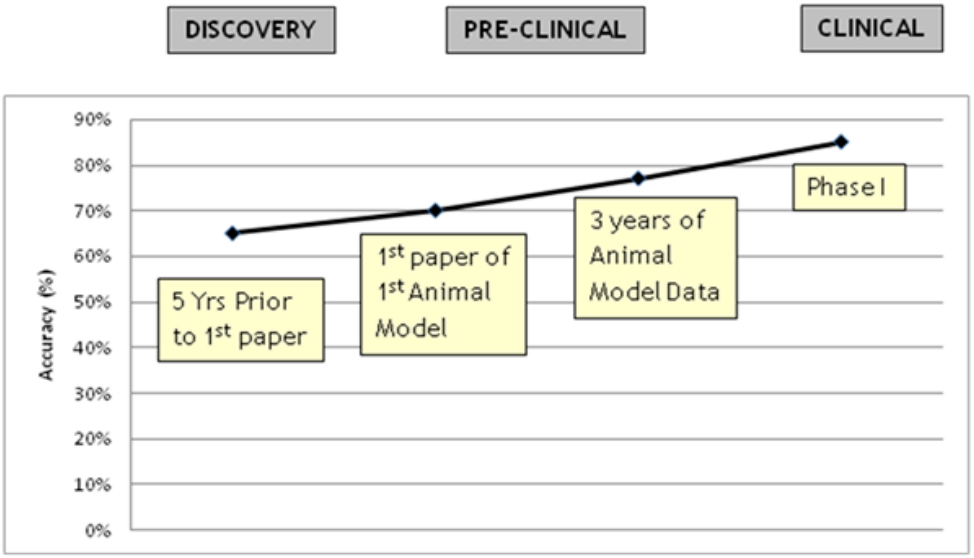
1. Man-machine interaction is paramount: it was decided that the human expert should not be taken out of the discovery loop but on the contrary represent an integral part of the discovery process. This meant that the system developed would not be expected to act as a black box that would deliver a single correct answer. Rather it should be flexible, very high performance so as to support real time interaction with the end user, handle the very large amounts of data in existence today and be truly scalable.
2. It should have in-built as few assumptions as possible: often systems that are used in a predictive capacity, namely mathematical models, have in-built our best understanding of the workings of the system whose behavior we wish to be able to predict, usually in the form of a mathematical or other model. In cases where this understanding is good and complete enough this can yield well performing predictors. In problems however where our understanding is not as complete, as is the case with biological processes and disease mechanisms, complex models are not as accurate and hence their predictive value diminishes. With this in mind, the technology employed is based on the simplest possible correlation between concepts and parameters of interest and relevance to problem solving, namely co-occurrence in a scientific publication. In other words we have made the simple assumption that if there exists a scientific publication that co-mentions any two concepts (for example a gene and a disease) this should be taken to mean that there *may* be some true biologically relevant correlation of these in a certain context (usually the context of the work reported in the paper). This approach is different from the ones that look to extract only those correlations that are deemed relevant and useful by some criterion and build from these their predictive models. Examples here are systems based on some mathematical model of a biological system or process as well as ontology-based systems.
3. It should return answers, not sources: to date Google is the quintessential search engine, but for the type of discovery we had in mind, returning 'relevant pages' was not what was required. We wished for the platform to return answers to specific questions (such as for example which set of genes could possibly act as a biomarker for a disease) rather than scientific papers that if read could possibly reveal the answer. On the other hand the platform should also support the user in understanding the answers generated, and in our case this was done by providing 'drill down' to the underlying literature.
4. It should rank answers: Discovery is often about choices and selecting amongst viable options. While in the end only one option is usually selected as the answer/solution the existence of the options and their eventual rejection contributes significantly to the better understanding of the solution and ones confidence in that solution. Therefore ranking of multiple answers was considered a significant performance parameter.

5. It should handle dirty and incomplete information: many platforms perform less well when the input data does not meet certain criteria such as completeness, relevance and being non-contradictory. Scientific publications however and knowledge in the case of biology are all of the above. Proponents of other approaches, point to the fact that if one does not perform any kind of a pre-selection, then one is subject to a lot of 'noise' as well as contradicting information such as for example that a gene A up-regulates *and* down-regulates the expression of some protein B. While we agree with such remarks, especially in view of the fact that knowledge of biology is continuously evolving and that what was previously thought correct can later be shown to be wrong, our assumption has been that a 'large numbers' approach where statistical significance starts to have effect, these contradictions are eventually taken care of and 'smoothed out'. This however remained to be proven and was one of the goals of our assessment of the technology platform reported in this document.

Assessing the Literature Mining and Discovery Platform

Before any predictive tool can be usefully employed, one needs to have an understanding of its predictive accuracy and how this may be affected by various parameters. For the assessment of the Biovista platform we set out to measure the following:

The predictive accuracy of the platform: We set out to measure the percentage of accurate predictions of the platform in the context of predicting adverse events of drugs as these are reported in the ASCO meetings. The test was a retrospective prediction test, meaning that the platform was made to work with historic data and checked against adverse events (AEs) reported in subsequent literature. The figure below shows the results obtained



The analysis showed that the predictive accuracy ranged from between 65% ad 85%. While the lower figure seems little better than random (50%) it should be compared with industry standards, which are in the very low 2 digits.

How the predictive accuracy varies over time: we were interested to determine how the predictive accuracy of the platform varies as a function of the amount of data available to it. Biovista's platform uses Medline abstracts; for the evaluation, the database of Medline abstracts was reduced so as to contain all abstracts from the beginning and up to specific years (1997, 2000, 2002, 2005 and 2007) and predictions using each of these five abstract

sets were checked against AEs reported in the 2007 ASCO conference. Using this approach, the useful predictive window was found to be up to a maximum of 7 years, meaning that the platform could reliably predict an AE up to 7 years prior to its first mention in the literature. Additional information and findings are reported in the paper submitted to Nature Biotechnology which is currently at the revision stage (see “Pro-active drug safety: combining existing data in new ways to predict serious adverse events of drugs.” S. Deftereos et al).

Use in the filed: we wished to determine how easily and effectively the technology could be employed in a real working environment. To determine this, we gathered usage data from actual users of this technology. We found 3 things:

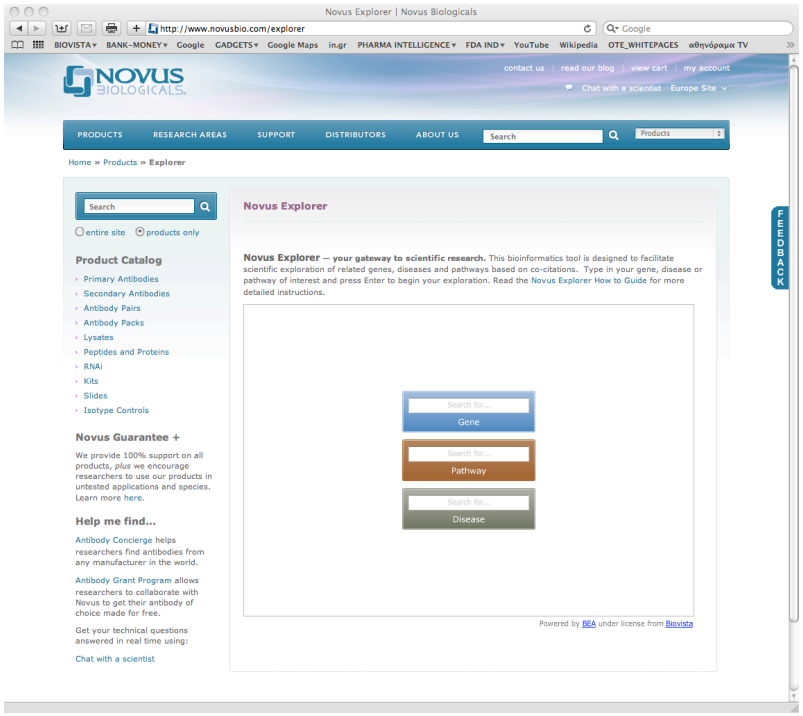
1. once understood, users found the technology useful
2. there is a certain kind of user that is best suited to such tools: this is a person who ‘likes’ discovery work, is open to IT tools and is good at making conceptual and scientific connections between seemingly unrelated ideas, possibly from disparate fields of expertise.
3. Use of IT tools and literature resources is not a mainstream activity of bio-researchers and clinical doctors.

Literature Mining for Sales support

Making the literature mining search algorithms available as a service, has allowed Biovista to provide solutions to third parties that were interested in sales support. One such case at hand is the Novus Explorer application developed for Novus Biologicals in the US.

Novus Biologicals (NB) is a reagent company that was interested in boosting its on-line sales via novel services that would add true value to its site visitors thereby increasing the chances that a sale would be made. Biovista’s literature mining capabilities were selected to power its *Novus Explorer* service. *Novus Explorer* allows visitors to NB’s site to explore genes and pathways related to an original gene or protein of interest, providing in addition the relevant literature that allows them to explore such correlations in more depth.

The figure below shows *Novus Explorer* on NB’s site (http://www.novusbio.com/explorer) ad the acknowledgement to Biovista’s BEA literature mining technology.



As a result of the Biovista technology, NB has seen a doubling of its on-line sales, already from the first year of the release of *Novus Explorer*.

Combining Literature with Patient Records - Thalassaemia International Foundation (TIF):

TIF is an international foundation that supports research, awareness and educational activities related to beta-thalassaemia. TIF was interested in deploying a patient record system at 3-8 collaborating hospitals in Cyprus (to start with) and later on in other countries with which it has ties. The management of patient records is along the lines envisaged by ACGT. Following a number of presentations and meetings, TIF agreed to install a specific EHR system (JAnaemia) in 3 hospitals in Cyprus. The system which has now been successfully installed is up and running and will be at the heart of a study regarding the care of beta-thalassaemia patients. This study will be presented in May 2011 at the 12th International Conference on Thalassaemia and the Hemoglobinopathies. The study will also see how patient record data can be combined with scientific literature analysis in order to achieve more personalized health care and can provide the basis for a collaboration with the p-medicine project.

3.6 CUSTODIX

Custodix's exploitation has focused on the CAT (Custodix Anonymization Tool) platform. CAT is considered a proof-of-concept for a generic de-identification tool, which allows to standardise the otherwise often ad-hoc pseudonymisation process. This standardisation is important in view of data privacy compliance and audit.

CAT will be made available as part of a service-oriented solution for a large NHS (National Health Service) project in the oncology domain. The work for which Custodix has been subcontracted will be based to a large degree on the CAT engine, which was initially developed during the ACGT project. Thanks to the pluggable architecture of CAT, the core functionality will be reused in the commercial service and will be extended by implementing plug-ins that will help addressing the specific customer needs. The CAT configuration tool (GUI front-end for configuring the CAT privacy profiles and for invoking the engine) will continue to be used at least for internal purposes (e.g. drafting privacy profiles).

The service-oriented solution will be further developed and exploited subsequent to this commercial project.

3.7 FORTH

FORTH has been the main developer of the Workflow Editor and Enactment Environment (WEEE).

The ACGT Workflow Environment offers an open web based platform for the design, management and sharing of complex scientific analyses. It can used to construct, perform, and share domain neutral scientific experiments, especially those that are data and computation intensive. These features are supported by the (optional) deployment in high computational infrastructures such as the Grid or in "Service as a Platform" emerging solutions e.g. Cloud Computing environments. FORTH therefore envisions that the benefits offered by the ACGT Workflow management system guarantee its sustainability and future enhancement and evolution beyond the end of the ACGT project. The plan for its future exploitation includes the following:

- In-Silico experimentation and modeling. The computer based modeling of the living organisms and the simulation of their function present a new domain where data integration and computational tooling and methodologies play an important role.

Already the use of the ACGT Workflow Environment has been decided in the context of new European projects such as the TUMOR and the p-medicine.

- Implementation of novel experiments in clinical and biomedical research. The use of the ACGT technologies in the Gen2Phen project for the realization of the methodology that enables the discovery of genotype-to-phenotype associations and predictive models based on SNP data is one such example.
- Community building and outreach. Delivering the whole environment as Open source software will help the building of a healthy community of users that through their active membership and feedback will further extend its functionality, usefulness, and exploitation. This plan nevertheless requires a lot of work in terms of the documentation of the code, provision of a development web site, email lists for developers and users, and so on and therefore requires careful planning and preparation.
- Future research. As a complex information system the ACGT workflow environment offers a large array of possibilities for state of the art research. Possible research directions include: the efficient execution of workflows into high performance computational platforms, high level workflow languages, the semantic and syntactic data integration, interoperability and integration of semantic web services and semantic workflows, semantic provenance and linking of the workflows and their results to the domain level user objectives, intelligent content delivery, personalization, and sharing, modern web based user interfaces, etc.

3.8 USAAR

The University of Saarland has pursued a vigorous exploitation of ACGT in a number of follow on projects and initiatives, including StaRC, Obtima, p-medicine and others which are already covered in other sections of this report.

ENCCA is another project in which USAAR and other ACGT consortium members are to exploit tools and technologies developed in ACGT. ENCCA aims to speed up and facilitate the introduction of safe and effective innovative therapies in the care of children and adolescents with cancer.

The ultimate aim of the ENCCA network is to

- increase cure for children and adolescents with cancer
- improve quality of life during treatment and quality of survival
- improve access to best standards of care and to increase the capacity to deliver this standard throughout Europe

Appendix C provides additional details on the ENCCA proposal.

4 Lessons Learned

The exploitation of a broad and complex EC collaborative project is a multi-dimensional problem with many challenges as well as opportunities. The three distinct frameworks of exploitation that were defined at the start of the project have provided the necessary context for organizing exploitation activities in a coherent manner. Nevertheless, not all hoped-for goals have been achieved. Even in these instances however, the experience has provided lessons that could provide additional guidance to others who might decide to work in the broad areas covered by ACGT.

In this section we present these lessons learned covering the development and exploitation of specific modules (like the Oncosimulator or the CAT module) as well as broader aspects of the project as a whole. Building on previous versions of this plan, we also revisit risk factors identified, report on which of those actually materialised and comment appropriately.

4.1 Approach and Deployment Specific Lessons

Selection of Middleware

The ACGT infrastructure is a complex technical infrastructure to which many different partners have contributed. Such large technical collaborations can only succeed when sufficient attention is given to using as much as possible industry standards. ACGT has done this; however, the infrastructure produced is still rather monolithical due to the use of its middleware (GLOBUS, GRIDGE). The big advantage of relying on the existing middleware is that it provides standard solutions to common complex tasks (e.g. delegation of rights, orchestration, resource allocation, etc). In hindsight, a technological undertaking of the scale of ACGT (with so many contributors) might benefit from a more lightweight solution. This means that one is burdened with a number of tasks otherwise covered by the middleware, but experience has shown that not all functionality offered by the complex middleware is strictly necessary in the trial application domain.

On the other hand a lightweight architecture can evolve more dynamically (e.g. adopting new technology) over time. Cooperation with a large number of people can also be more efficient if the focus of collaboration is on interfacing, interoperability and integration in a more loosely coupled architecture. Quality control and acceptance testing in view of compliance could also benefit from this approach.

It should also be noted, that although a more lightweight (loosely coupled) architectural approach can in practice easily evolve into “a new” heavyweight middleware (as more functionality is shifted to centrally maintained and deployed proprietary components), the potential reward makes the approach worth investigating.

Deployment

In terms of deploying the ACGT infrastructure and resources in end-user environments consortium exploitation efforts have highlighted a number of issues that merit specific attention by similar efforts in the future. They are the following:

1. *Not enough infrastructure and services in place*: the lack of enough related end-user oriented modules in ACGT is believed to be one of the factors that have hindered its deployment in actual working environments. In particular, the ‘completeness’ and consistency of the available set of end user services has not been what is required to support existing workflows and tasks in end user partner environments (like Jules Bordet, University Hospital Saarland and others). While considerable effort was spent in the last year to define scenaria that represent actual tasks and workflows that make

sense to end users (doctors, researchers, etc.). these were still 'fragmented' being able to support only simple operations.

2. *Opt-in versus Opt-out and the anti spam laws*: The ACGT MB decided to adopt a conservative approach in all its mailing and communicatios efforts. Opt-in rather than opt-out policies were followed. As a consequence, circulation and site visit figures were rather low. It is felt that the failure of the ACGT competition to attract sufficient interest was also partly due to this policy which should not be adopted by similar projects.
3. *Continued support of ACGT infrastructure*: A number of contacted third parties expressed this concern at various points in time. While we are not able to ascertain to what degree this has acted as a true deterrent to adoption, this is certainly an important concern that must be addressed before successful uptake can be expected. The success of ACGT related project proposals (such as p-medicine) will provide some continuity but once again the p-medicine consortium will again be faced with such a concern that it will have to address early on.
4. *Intellectual property*: IP-issues tend to be a hindering factor in data exchange: On the one hand participating clinicians have serious reservations against sharing (raw) patient data as their possession is an important (and not always legally protected) factor in scientific competition. On the other hand patients' (sometimes economic) interests in the outcome of the research are not always sufficiently covered by trial setups and results' exploitation. ACGT has developed guidelines for bringing decision makers into the position to allow patients and clinicians proper participation in the exploitation process. This work can serve as a basis for European project managers in the E-health area to identify intellectual property issues in an early stage of the project's lifecycle.

4.2 Module Specific Lessons

Central Authority for Data protection - Custodix

One clear conclusion from the ACGT work is the need for an independent central authority. This need has been established in the beginning of the project as a requirement following from the legal analysis (cf. D10.2 "The ACGT ethical and legal requirements"). Within ACGT it has lead to the establishment of the "Center for Data Protection" (a non-profit organisation under Belgian law) that can act as data protection authority in international cooperation on medical research. This conclusion that was made early on in the project has only been further reinforced during over time (note that other projects such as caBIG independently came to the same conclusion).

We have learned that in order to come to a sustainable operational1 infrastructure, such an authority should be attributed sufficient power. ACGT as a research project was never intended to deliver a commercial-grade operational solution. Still, already during the pilots, many difficulties of managing (mainly enforcing) a stringent legal and security framework (as required) have been experienced. These issues can only be thoroughly dealt with if a central authority is in place which is enabled and empowered to check (audit) for non-compliance (both administratively and technically) and take appropriate measures against violation, such as monetary penalties and exclusion of partners (i.e. can police the collaborative environment).

[1] Operational not in the sense of "technically working", but in a broad sense "running in exploitation".

In a sense, this goes beyond the scope of a research project, but we have experienced during ACGT that sustainability (especially when this implies serving the commercial pharmaceutical industry) can only be reached with a sufficiently powerful authority in place. Technical tasks of such an authority (off course such tasks can be delegated) could include a compliance check whenever new services are started, but should at least include regular technical on site audits at the sites of all organisations participating in the infrastructure.

Thus, an important conclusion from the ACGT experience is that in order to create an exploitable platform, one must install and sufficiently empower a central authority and foresee the necessary operational funds in the business planning stage for enabling it to perform its duties.

Compliance with legislation - Custodix

Compliance to trial related legislation, especially to the data protection laws, is a critical success factor for any research-network. ACGT has expended considerable effort in order to automate achieving this compliance, for example through the founding of CDP (for establishing necessary contracts) and with tools such as CAT (Custodix Anonymisation Tool).

ACGT has shown us, that investing in achieving compliance by default (without specific effort or much expertise from the end-users) with fixed procedures is a must for long term success and smooth operations. In view of the experience, one could suggest that this requirement should be valued very high when making technological choices.

Literature mining for Systematic Discovery - Biovista

While the evaluation described above has focused on the platform itself, we believe that the findings apply equally well to the context of ACGT and in particular its goal as a supporting infrastructure for clinical trials design and management.

1. Finding 1: Literature mining and literature based discovery have been shown to work in a predictive capacity. Both can be used to create new knowledge from existing scientific publications and can therefore support research in biology and the development of more effective therapies in a clinical trials and drug development setting.
2. Finding 2: While the output is biologically and clinically relevant, IT tools are still not regarded by researchers and doctors as central to their profession. It therefore takes a 'special breed' of such practitioners to appreciate and effectively use these technologies. The main characteristics of such early adopters and subsequent 'champions' of these technologies include:
    a. Positive attitude towards the incorporation of new tools in ones daily procedures
    b. Aptitude towards IT tools in particular
    c. Interest in discovery and the ability to correlate seemingly disparate information
    d. Thorough understanding of at least one area of interest (e.g. CNS diseases) accompanied by a good general knowledge of other fields
3. Finding 3: Literature based discovery in the context of clinical trials design and support is not a primary concern. Data integration, patient record management and other such activities are more important. Discovery services have a role in investigator initiated clinical trials and areas such as for example biomarker discovery and treatment design which of course are themselves a part of clinical trials.
4. Finding 4: Since they have been shown to have a real utility, such services could be usefully employed to illustrate the broader utility of IT infrastructures. ACGT aims to

deliver an infrastructure in support of clinical trials. However, just like any resource it requires the end user to invest personal time and effort to learn and effectively use it. Successful services can provide apt justification for why such an investment might be worthwhile and should therefore be promoted in that light.

Developing Simulation tools – Oncosimulator by NTUA

A deeply-founded background on a host of diverse scientific, technological and medical domains is a sine qua non prerequisite for a fruitful involvement in the nascent domain of in-silico oncology, including the development of oncosimulators.

Since the identification, collection and legal exploitation of multiscale medical imaging, histological, molecular, clinical and treatment data needed for driving, clinically adapting and validating multiscale cancer models is an extremely demanding endeavor, related issues should be addressed and solved from the beginning of any such work.

The strongly interdisciplinary character of in silico oncology dictates an open minded and deeply collaborative approach to highly complex problems of clinical importance.

Since in-silico oncology research presupposes a continuous, complex and highly demanding interaction among scientists, engineers and medical doctors, a careful selection of partners as well as mutual understanding is of the utmost importance. Due to the fact that in-silico oncology is still under formation, and completely new ideas, approaches, methods and techniques have to be proposed and tested, a lot of patience and perseverance are needed.

Semantic Integration- UPM

Developing a semantic integration layer for biomedical databases resulted in a quite complex task. The biomedical domain, and more specifically, the cancer-related clinical trials domain, evolves at a surprisingly high rate. We found out that, during the four and a half years that the project lasted, new requirements appeared, or initial requirements had to suffer modifications, simply because the biomedical field had new needs. We found it was crucial to adopt highly flexible designs for our tools, so that they could be adapted to the new needs without requiring deep changes in the code.

Due to the extension and the rich feature nature of the tools to develop, it was necessary to adopt third party tools or APIs available for the research community—e.g. Globus framework [1], OGSADAI framework [2], Jena API [3], Google Web Toolkit [4]. Special care must be taken when selecting what tools to use, since lack of documentation or support can seriously affect development.

The Grid Environment

The proper set up and operation of the underlying technologies present a significant barrier to use and adoption of this technology. For example, experience gained during ACGT showed the following:

- No loose ends should be left during installation. Even date synchronisation between PCs is important (very useful NTPServers).
- Use complete sets of libraries. Significant dependences among different versions of

Globus Toolkit (wrong version of library may cause incorrect execution) exist.

- The curse of security. In the grid the security is very high, and so is the risk of failure. Attention should be paid to the process of generation and delegation of credentials.
- Middleware software.Very useful in this context (at least in the case ACGT project and PSNC tools).

4.3 Risk Factor Assessment

As part of its Exploitation Plan the ACGT consortium identified in the early stages a number of 'risk factors' that could affect the exploitation of project results. In this section we revisit these factors, reporting on the ones that materialised and providing comments and guidelines for future reference.

	Risk Factor	Comment	Lessons Learned
1	Early prototypes fail to raise interest	While interest has been raised, this was not sufficient to ensure further commitment and uptake	The large scale prototype planed for May 2008 was created and demonstrated at Review 2. The consortium is now concentrating on stability issues. At the same time these prototypes have been put in the hands of end users and at the consortium meeting in Vienna (January 2009) tutorial sessions were run for their benefit. This and similar actions are considered necessary to help end users get started on the 'ACGT learning curve'.
2	Use Case studies not convincing	• Work was carried out to ensure they represent clear, present and important needs of ACGT's targeted end user communities • The consortium also examined the requirements of the BIG initiative with the aim of providing a solution to a present issue they are addressing.	IT is of paramount importance for such large scale efforts to ensure dedicated end users with a clearly articulated need that will be eventually addressed by project work, has specific milestones during the project, ad has internal champions who will ensure uptake. The use cases must represent present, clear, true needs that will establish credibility of the new solution and allow transfer of higher value problems to the new solution.
3	Legal/ethical impediments to widespread use	While these exit, they were not an impediment for many of the goals of ACGT.	More work needs to be done to addressed the fragmentation at the EU level.

4	Competing initiatives achieve critical mass or backing by important stakeholder groups overshadowing ACGT	ACGT established links with other initiatives in the space (EGEE and caBIG) that will be exploited in follow on projects.	Consider competition, identify on uniqueness, and establish links where appropriate.
5	Awareness efforts lag	ACGT visibility by project end is satisfactory with consortium members having published and presented at a significant number of conferences and other fora.	The momentum should be kept up ensuring continued references to ACGT where appropriate in order to build on present achievements. Follow in projects should not 'start from 0' but clearly sow links that prove continuity and relevance of work.
6	Technological developments in the area of Grid services render ACGT options obsolete	There is already talk (e.g in the recent calls for EC proposals) of 'cloud computing' as the next thing to follow the 'grid'. It is still early days and at this point 'cloud computing' is only confusing potential end users.	The consortium believes that in such complex problem spaces, the winners will be those who offer useful services in a transparent, easy-to-use and well integrated manner.
7	Failure to convince 3rd parties s/w providers to contribute to ACGT	While sufficient materials and critical mass that would attract 3rd parties to develop ACGT compliant resources and services were developed, as evidenced by the cancelled ACGT competition these were not sufficient.	There is a need for a very targeted identification of 3rd parties with an important stake in the selected area, who also share at least part of the 'technical vision' of the project and will be willing to make the necessary commitment of resources. Such efforts must start as early as possible in order to establish contact, show continuity and prepare the ground for when the infrastructure is eventually ready to support 3rd party contributions.
8	Working environment and administrative complications hinder adoption	This occurred for ACGT and seems to be the case for any such complex offering.	Significant effort must be spent in the early days of such projects in identifying the appropriate large partners that will actually deploy such solutions and act as success stories that others might follow. Probably this criterion should be assessed at the time of proposal

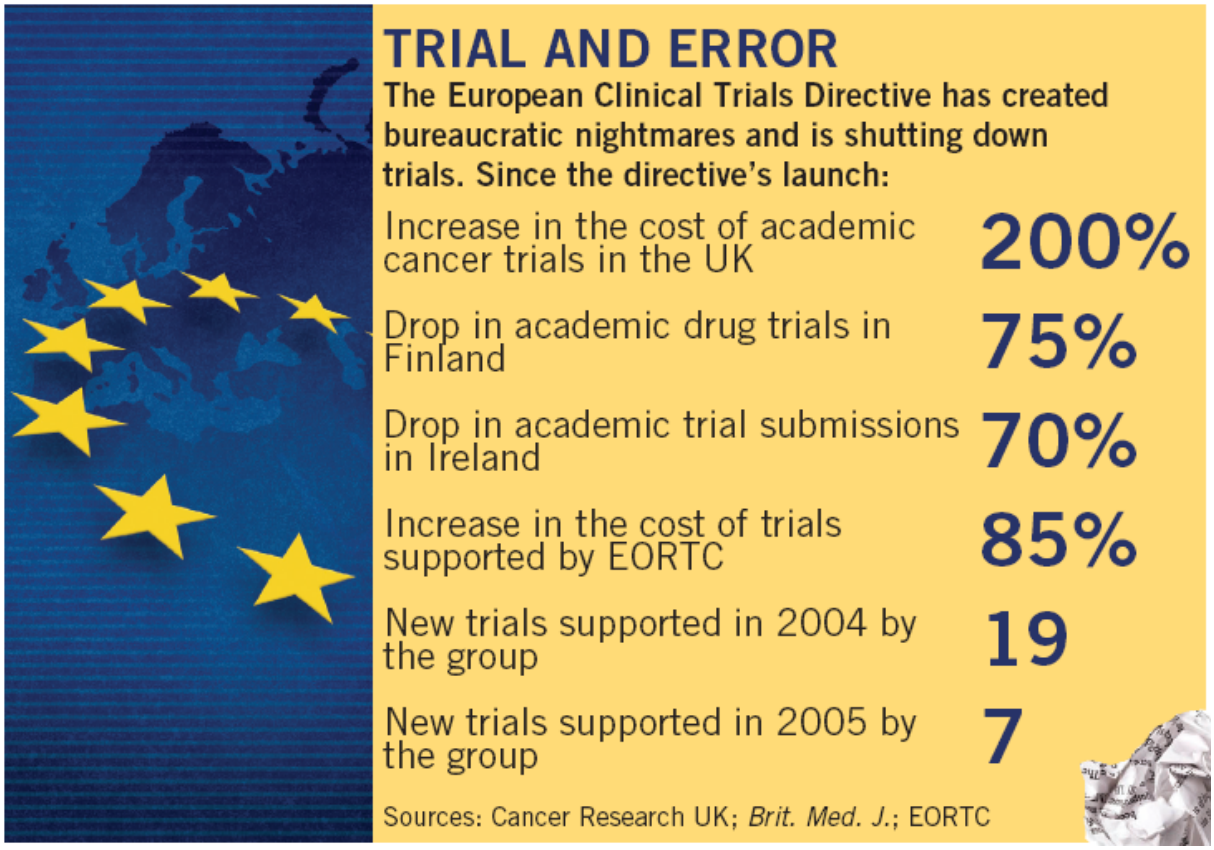
			evaluation rather than post proposal approval.
9	Legal and ethical considerations provide larger than expected restrictions	This has proved correct. Significant output has been generated by the consortium. It is felt however that at this point in time, it is not possible to deploy a European wide solution easily.	Legal and ethical considerations represent a multi-faceted problem that is hard to address at a EU level. It is felt that complete projects should be dedicated to exploring and pushing the boundaries in order to reach some commonly acceptable framework that will be workable.

APPENDICES

APPENDIX A. The STaRC (Study, Trial and Research Centre) Initiative

Clinical trials are essential to achieve better treatments for patients. As a result of the Clinical Trials Directive 2001/20/EC the conduct of clinical trials throughout Europe has changed (1, 2). The directive, aimed largely at holding pharmaceutical companies to higher standards, has tied up academic clinical research, particularly large trials, with redundant paperwork, liability tangles and unending bureaucracy (1).

Figure 1: The Impact of the European Clinical Trials Directive 2001/20/EC (1; figure taken from the article)



Brandon Keim writes in Nature Medicine: "The cost of academic cancer trials has doubled since 2004**,** according to Cancer Research UK, the country's largest sponsor of academic cancer research. The European Organization for the Research and Treatment of Cancer estimates that expenses have risen by 85% and says the number of trials it supports has dropped by 63%. The Save European Research campaign, which represents more than 3,000 scientists, says academic drug trials have dropped by 70% in Ireland and 25% in Sweden. The number of Finnish academic drug trials shrunk by 75%". (1). One of the biggest bottlenecks is the directive's requirement that each trial has to have a single sponsor who is fully liable for all legal and financial issues. For trials running in different European Countries the problem of a single sponsor is not solved yet. Key issues for Cancer Trials are summarized by Kathy Pritchard-Jones in the European Journal of Cancer (2). Though this

article deals with clinical trials for children, most of these points are relevant for clinical trials in adults.

Scenarios and structures that help to run more clinical trials and to bridge the gab between treatment given to patients today and research to find better treatment for patients is of utmost importance. This is the most relevant reason to build up **STaRC** - a Study, Trial and Research Centre – that will support translation research in all aspects.

Table 1: Key issues for Paediatric Cancer Trials in relation to the EU Clinical Trial Directive 2001/20/EC. (2)

Issue	Experience of European paediatric study groups running investigator-led ('non-commercial') trials in childhood cancers
Definition of an interventional clinical trial	'Standard of care' regimens often include medicines used 'off label' Variation in acceptance by national regulatory authorities of such use as 'background medicine' or whether it falls outside the definition of an 'interventional clinical trial'
Sponsorship	National variation in whether a single European sponsor is required or a national co-sponsorship arrangement is accepted Complex contractual negotiations required between partners
Insurance and Indemnity	Large variation in costs and in whether 'no fault' indemnity is required Insurance costs increased 100-fold with no perceptible change in risks between consecutive trials of the same study group Premiums may be paid by fundraising efforts of childhood cancer parents' associations
Definition of an IMP	Hugely variable for use of old drugs with no or limited paediatric information in their marketing authorisations IMP definition has major impact on bureaucracy of pharmaco-vigilance
Pharmaco-vigilance	Hugely bureaucratic with no noticeable improvement in patient safety (which was in any case very good in childhood cancer trials) National variation in onward reporting requirements for SUSARs when drug is used in more than one trial Inconsistency in inspection findings of regulatory processes for the same trial
Sponsor obligation to provide free drug	Large national variations in how this is absorbed into national health insurance schemes or whether this must be paid for by sponsor Required for IMPs, whose definition is also variable
Drug formulations adapted for children	Lack of appropriate formulations for young children for many oral anti-cancer drugs Strict definition of 'manufacturing' excludes young children from some clinical trials when no appropriate formulation exists
Ethical considerations	Ethical committees need appropriate expertise to evaluate appropriateness of new drug trials in children Timelines to receive the 'single' national ethical approval highly variable Institutions have created other hurdles to opening a trial, variably labelled 'R & D' approval

In detail the following problems in clinical care of patients do exist today:

- There is a time lack for physicians being kept informed about all the new developments in medicine, even in their specialized field. Every week hundreds of new papers are published. To find the most relevant, to read them all and to judge them as important for the own work is impossible.
- Today teamwork is of utmost importance. No physician is able to treat a patient with cancer by his own. He always has to communicate and work together with other specialists in medicine. As a result a lot of so called Cancer Comprehensive Centres are established to facilitate the interdisciplinary work. But up to now no IT

infrastructure is supporting this by storing all relevant data in a database, so that every treating physician will have immediate access to the history, diagnosis, treatment and other relevant data of patients in an anonymous and secure way.

- Physicians do not get feedback of how efficient they are working. They do not have any statistics regarding the survival of their patients compared to the survival of all patients with that kind of cancer. There is no benchmarking telling them they are doing good or bad.
- Physicians do not know about the possibilities of modern IT technologies that could help them to support them in daily care of patients, or in developing new clinical trials. The lack of this knowledge leads to a lack of requests and requirements to IT people for the creation of new and user friendly tools in this respect.
- Only a minority of patients are enrolled in prospective clinical trials. The reason for this is manifold:
    - Physicians do not (want to) enter patients in clinical trials because
        - they fear the burden of workload by entering patients (documentation, regulatory and administrative necessities, etc.)
        - they are not well informed about the meaning and impact of clinical trials (fear of experiments with their patients, simply not used to enrol patients in clinical trials, etc.)
        - in most curricula of Medical Schools Clinical trials are missing, so that students will not learn about the beneficiaries of clinical trials
    - Patients do not want to enter a clinical trial
        - they are not informed at all about clinical trials
        - they are not well informed about the meaning and impact of clinical trials (fear of taking part in an experiment, etc.)
    - There is no financial and/or administrative support to cover the overhead of clinical trials
        - the burden of European regulations contrasts the available resources to increase the number of new clinical trials
        - infrastructures in hospitals or outpatient facilities are lacking (no data manager, etc.)
- There exists no database for clinical trials with an easy way of access for physicians or patients in Europe. This is twofold unacceptable:
    - A physician is not able to find the best trial that fits the need for his patient
    - A trial chairman might build a new trial that is still running by another physician
- Today patients do use the internet to get information about their disease. There is no way how a patient can trust such information. Often information is contrary and alienates patients.
- Even if patients do find relevant information, they may not understand the medical language used in these information.
- More patients are asking for second opinions regarding their disease. This is time consuming for physicians, expensive for the health care system and often unsatisfying for patients. They often get different and contrary answers resulting in the question: “And what should I do now?”

A part of these problems are already faced in ACGT (Advancing Clinico-Genomic Trials) an Integrated Project that is funded by the EU. Norbert Graf is the only clinician in this project, he is the quality manager of ACGT and the leader of work package 2 of this research project. ACGT has already built up an IT infrastructure to run clinico-genomic trials throughout Europe. A data security framework for clinical trials exists via the 'Center of Data Protection' (CDP; https://cdp.custodix.com/) and software tools like CAT (Custodix Anonymization Tool) that are developed within ACGT. Custodix is a private limited company established in 2000 specialized in data protection solutions for eHealth (https://www.custodix.com/) and recognized as one of the most advanced and reliable Trusted Service Providers (TSPs) in the Healthcare sector. The main problem of ACGT is maintenance and sustainability after the end of the funding period by the EU. The perspective of **STaRC** will be to exploit but not perpetuate ACGT. Both the CDP as Custodix are willing to participate in STaRC.

Figure A1: Logo of STaRC

The main tasks of STARC are:

1. Simplification of the clinical trial process
2. Patient empowerment
3. Combining clinical and molecular biological and genomic data in single patients leading to personalized medicine
4. Facilitating translational research
5. Continually improving curricula of medical schools and medical education

The structure of **STaRC** today is a CRO (Contract Research Organization) that can easily work together with other Organizations, Registries and external Centres. STaRC is more than a Comprehensive Cancer Centre (CCC), because of the following reasons:

- it will deal with patients having cancer and all other kind of diseases
- it is also a research organization fostering translational research
- it has educational and teaching aspects
- it links patient care with research and teaching
- it has an IT infrastructure

- it has a structure to provide help for patients throughout centres and countries within Europe
- it provides a security framework for patients
- it empowers patients
- other organizations and pharmaceutical companies can cooperate

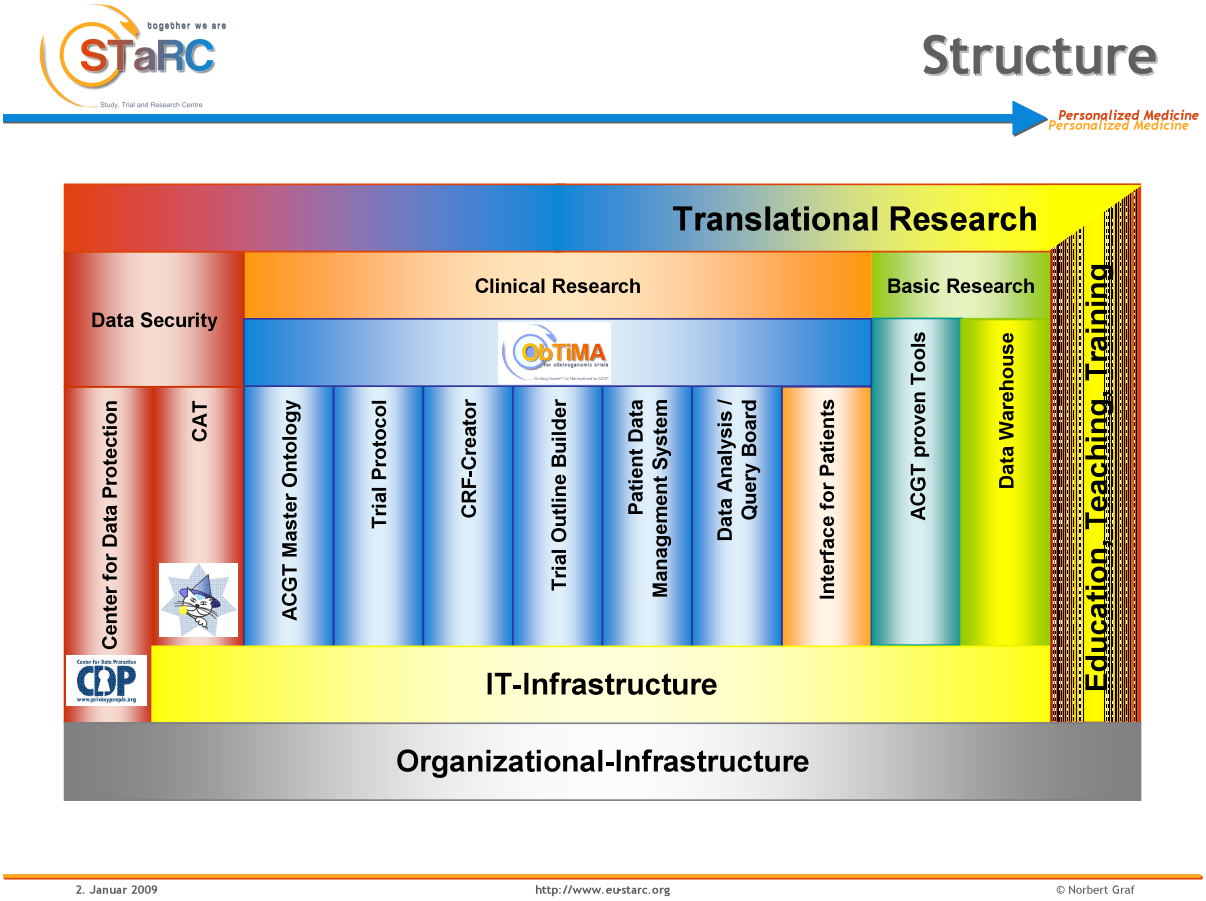


Figure A2: the Structure of STaRC

The organizational infrastructure of STaRC is shown in Figure A3 below.

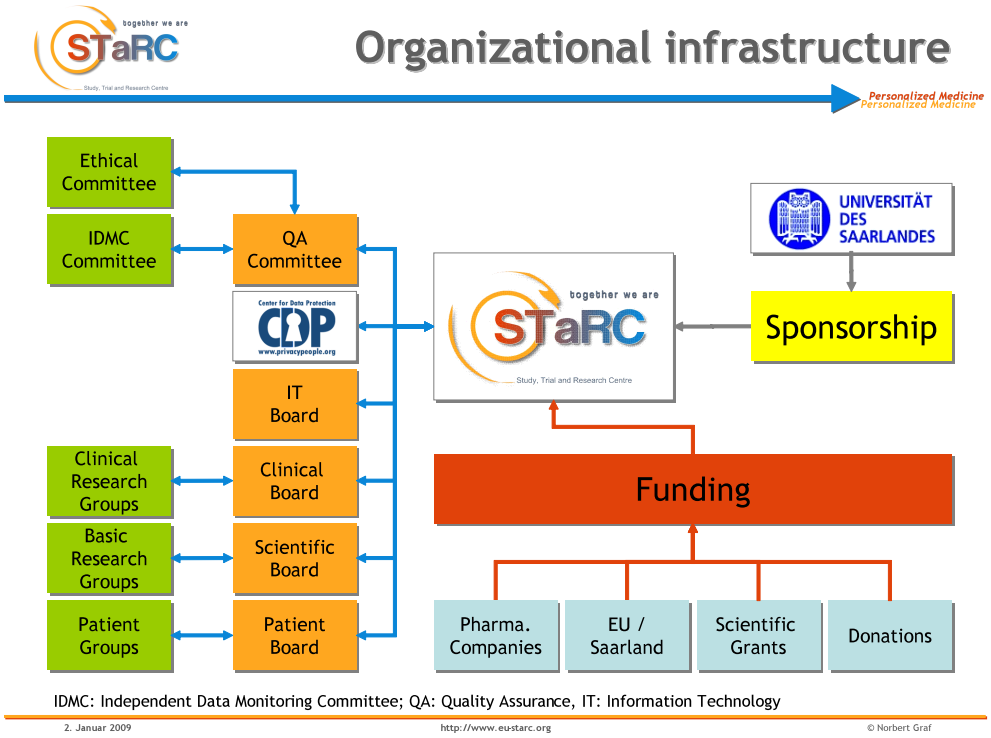


Figure A3: The organizational infrastructure of STaRC

One of the most important software tools for running clinical trials is ObTiMA (Ontology based Trial Management Application). ObTiMA is an open source software developed within ACGT by different partners based on the idea of Norbert Graf:

- Fraunhofer, IBMT in St. Ingbert
- Foundation for Research and Technology Hellas (FORTH), Institute of Computer Science, Vassilika Vouton P.O Box 1385, GR-71110 Heraklion, Crete, Greece
- Meme Media Laboratory, Hokkaido University N-13, W-8 Sapporo, 060-8628 Japan

This software will be used in STaRC to run clinical trials. The functionality of ObTiMA is manifold as shown in figure 3. ObTiMA is a modular tool having the following subcomponents:

- A tool for writing a trial protocol by being guided through all legal and ethical requirements
- A Case Report Form (CRF) creator
- A Trial Outline Builder (TOB) as a graphical interface to a trial
- A Patient Data Management System
- A Data Analysis and Query Board
- An Interface to patients

ObTiMA can be used with and without an Ontology. The use of an Ontology will allow cross trial analysis to gain better and faster trial results without performing metaanalysis.

This proposal needs to be discussed to find the best infrastructure for starting **STaRC**. There is a great chance to develop **STaRC** in a way that it will get the main Centre for running clinical Trials within Europe. As Norbert Graf is one of the clinicians in ACGT, and as there is an enormous interest of the EU to maintain and sustain the results and advantages of ACGT this chance is a realistic one. Other partners that should be involved like Custodix, the 'Center for Data Protection' and the Meme Media Institute at the University of Hokkaido,

Japan are agreeing to build **STaRC** in the Saarland area. A close connection to the University of the Saarland is attempted by the Medical Faculty.

Partners that are involved in **STaRC** are:

- University of the Saarland (Medical Faculty (Dep. Paediatric Oncology), IFOMIS, Fraunhofer (IBMT))
- University of Hannover (Lehrstuhl für Rechtsinformatik und IT-Recht)
- Biovista, Athens, Greece
- Custodix, Brussels, Belgium
- FORTH, Heraklion, Greece
- University of Hokkaido, Japan, UoH

STaRC will be lead by the Dep. of Paediatric Oncology and Haematology. In **STaRC** three main topics are identified in for maintaining ACGT. These topics are

- Service
- Research and Development (R&D)
- Education, Teaching

The STaRC initiative has not taken a format structure yet, but both regional resources from the local government of Saarland as well as from patient support groups have been assured. Its formal establishment is a matter of time.

References

1. Keim B: Tied up in red tape, European trials shut down. Nature Medicine 13:110, 2007
2. Pritchard-Jones K: Clinical trials for children with cancer in Europe – Still a long way from harmonisation: A report from SIOP Europe. European Journal of Cancer 44:2106-2111, 2008

APPENDIX B: The p-medicine project Proposal Abstract

Medicine is currently undergoing a major revolution that is gradually transforming the nature of healthcare from *reactive to preventive*. The changes are catalyzed by a new systems approach to disease that have triggered the emergence of personalized medicine — a medicine that focuses on the integrated diagnosis, treatment and prevention of disease in individual patients.

This change is rooted in *new science*. The convergence of systems approaches to disease, new measurement, modeling and visualization technologies, and new computational and mathematical tools are expected to allow our current, largely reactive mode of medicine to be replaced over the next 10 to 20 years by a personalized, predictive, preventive, and participatory medicine[2]. While the goal is clear, the path to such advances has been fraught with roadblocks in terms of technical, scientific, and sociological challenges.

P-medicine brings together internationally recognised leaders in their respective fields with the aim to create an innovative computational, service-oriented infrastructure that will facilitate this gradual translation from current medical practices to personalized medicine. In achieving this objective *p-medicine* has formulated a coherent, integrated workplan for the design, development, integration and validation of all technologically challenging areas of work.

Our emphasis in on drafting an open and modular architectural framework for the tools and services to be developed, so that adoption of the p-medicine services will not be an all-or-nothing decision; on efficient sharing and handling of the enormous personalized data sets — including policies, security, modeling, cloud storage, etc.; on enabling demanding Virtual Physiological Human (VPH) multiscale simulations, for which standardization and semantic data integration and interoperability is a major issue addressed; on building and standardizing tools and models for VPH research, such as the VPH Toolkit[3], by defining a formalism to make the knowledge that is implicitly encoded in these tools explicit and thus improve the re-use of tools and solutions; on providing tools for large-scale, privacy-preserving data mining, and literature mining, a key factor in VPH research. On the policy front, we focus in making sure that policies with respect to privacy, non-discrimination, and access are aligned to maximize both the protections and the benefits to patients.

The p-medicine tools and technologies will be validated in concrete setting of advanced clinical research. Pilot clinical trials have been selected based on the presence of clear research objectives, raising the need to integrate large multilevel datasets, requiring innovative collaborative tools and technological support of highly distributed research groups, in the domains of a) a Wilms tumour trial testing the newly developed and validated tools of *p-medicine*, b) a breast cancer neoadjuvant pharmacodynamic phase II trial that will be used to extend current VPH tools and c) a leukaemia trial that will be used to run VPH models predicting minimal residual disease and recurrence in childhood acute lymphoblastic leukaemia.

[2] http://www.cra.org/ccc/initiatives

[3] http://www.vph-noe.eu/wp3

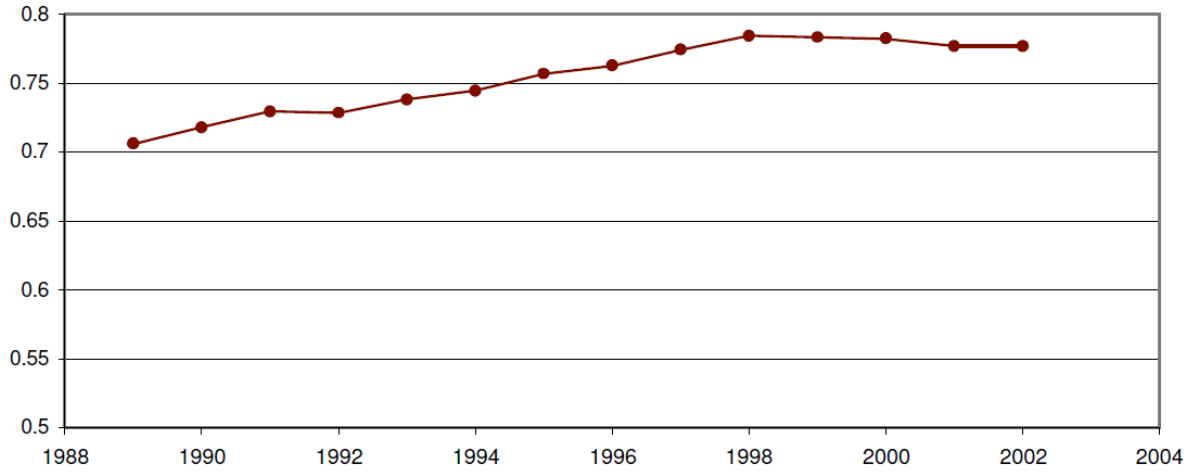
APPENDIX C: ENCCA (European Network for Cancer research in Children and Adolescents)

Concept and objectives

Cancer is rare in children and has a distinct 'embryonal' biology that is very different from the typical epithelial cancers of adulthood. Cancer affects approximately 1 in 500 during childhood and adolescence. This represents only 1% of all cancers in humans, yet remains a significant health burden for our young citizens. While overall survival rates have doubled from 40% in the early 1970s to nearly 80% today (Gatta, 2009), cancer remains the leading cause of death from disease beyond infancy in Europe (Pritchard-Jones K, Eur. J. Cancer, 2006). Almost 20,000 young people (aged up to-19 yrs) will be diagnosed with cancer this year in the EU. The majority can expect to be cured but survival rates are disproportionate across Europe and there are significant health problems in many survivors. This is a consequence of the long-term side effects of many current chemotherapy drugs and radiotherapy administered to the growing child. It is estimated that there are currently between 300,000 -500,000 adults in Europe will have survived a cancer treated in their youth. This number is set to grow as survival rates improve.

The improvement in survival during the last 40 years has been achieved through mainly national and international collaborative clinical trial groups setting the standards of care for these rare diseases and constantly seeking to improve outcomes through investigator-driven clinical trials (IDCT). To continue to improve survival rates, there is an increasing need for wider collaboration in order to enrol sufficient patients for properly powered studies. To date, the necessary multinational collaboration has developed on an 'ad hoc' basis, often dictated by geography, shared language or cultural values. The recent increase in resources and infrastructure required to IDCTs has placed further strains on these informal networks and highlighted the need for better integration and shared procedures. At the same time, it appears that progress in improving survival rates using current conventional chemotherapy agents has reached its limits and new approaches as well as new targeted therapies are needed.

Figure 1: 5 year overall survival, all cancer, Europe[4],[5]



[4] EUROCARE 4 (1995-2002)

[5] Gatta G et al, EJC 2009

Despite the high cure rates achieved with current multimodality treatments, cancer remains the leading cause of death due to disease in children over the age of one year. The challenge of the next decade will be to introduce the new generation of biologically targeted drugs to continue to make progress in improving survival rates and to improve the quality of cure of long-term survivors of childhood cancers. Safe and effective innovative therapies are urgently needed to improve cure rates and the quality of cure in children. Advances in knowledge of the human genome, cancer cell biology and the development of high-throughput technologies have increased our knowledge of tumour biology. This has paved the way for identifying targets for the design of anticancer compounds with new mechanisms of action, new profiles of antitumour activity as well as new toxicity profiles. There has therefore been a dramatic increase in the number of new anticancer compounds under development worldwide, and oncology has become the leading area for drug development. However, the 'translational gap' between the basic scientific knowledge of paediatric tumour cell biology and defining the mechanisms that 'drive' these cancers remains wide. Furthermore, access for children to innovative compounds developed for adults by pharmaceutical companies has been extremely poor in Europe in the last 20 years. This is in contrast to the USA, where many public programmes have provided easier access to new compounds to the paediatric oncology community. ENCCA will speed up and facilitate the introduction of safe and effective innovative therapies in the care of children and adolescents with cancer.

More precisely **ENCCA** aims to:

1. Structure and integrate on a sustainable basis, and European scale, the way that clinical trials in children and young people with cancer operate by proposing a **European Strategy for Clinical Trials** in Paediatric and adolescent Oncology
2. Use **efficiently** the existing European research **tools and equipment** in paediatric oncology.
3. Promote a **common methodology for clinical trial** design, implementation and integration
4. Initiate **harmonised therapeutic strategies by** increasing significantly access to knowledge about paediatric tumour biology and interactions in between tumour and host. Facilitate strategic discussions and joint research between scientists and clinical investigators for prioritisation of drugs to be studied in the paediatric age range and the translation of this knowledge into more personalised medicine for children and adolescents with cancer
5. Facilitate a **wider sharing of knowledge** and technologies across disciplines and the chain of all stakeholders in Europe (Academia, Parents/Patients organisations, Pharmaceutical companies and Regulatory bodies) and improve the **career structure in paediatric and adolescent oncology clinical and translational medicine.**
6. Improve substantially the quality of life of children and adolescents with cancer.
7. Propose common ethical definitions of issues and solutions adapted to national and cultural requirements

The establishment in parallel of a permanent **European Clinical Platform or Council for Paediatric and Adolescent Oncology** would:

- Contribute significantly to the political vision of creating a European Research Area, including a European Clinical Paediatric and Adolescent Oncology Area.
- lead to a more effective and cost efficient use of Clinical trial Tumour groups' expertise and resources
- facilitate decisions on future research infrastructure requirements and investments.
- better inform the evolving EU policies for clinical trials and medicines for children.

The ultimate aim of the ENCCA network is to

- increase cure for children and adolescents with cancer
- improve quality of life during treatment and quality of survival
- improve access to best standards of care and to increase the capacity to deliver this standard throughout Europe

FORTH is partner in ENCCA and works in several work packages. USAAR is associated partner in WP 2.4 (task 2.4.3). In WP1.2 (Establishment of the virtual institute information portal) FORTH is leading task 1.2.3 (Analysis of existing systems and solutions) and enrolled in WP 1.2.4 (Definition and management of standards and workflows), WP 1.2.5 (Design and specification of the data exchange and interoperability infrastructure), WP 1.2.6 (Development and setup of a infrastructure prototype) and WP 1.2.7 (Pilot operation and system validation). In WP 1.3 (Clinical trial facilitation) FORTH is enrolled in task 1.3.3 (Improving the framework for IDCT). In WP 1.4 (Biology to guide innovative targeted therapy development) and WP 1.5 (Standardised and innovative methodology for clinical trial design and analysis) FORTH is enrolled in task 1.5.4 (Definition of disease specific core data sets for pooled analysis). In WP 1.6 (Development, implementation and Evaluation of tools for non-invasive, functional imaging) FORTH is enrolled in task 1.6.3 Incorporation of functional imaging into clinical trials protocols. IN WP 2.2 (Improved therapeutic strategies using predictive biomarkers in leukaemias) FORTH is enrolled in task 2.2.2 (Establishment of a harmonised pipeline for molecular diagnostics in a European virtual laboratory setting using very high-risk ALL (VHRL) as a model system), task 2.2.3 (Integration of a molecular diagnostic pipeline with preclinical model systems for molecular targeted treatment and application of algorithms for identification and priorisation of molecular targets) and 2.2.4 (Harmonization and integration of clinical platforms for the introduction of molecularly targeted treatment in leukaemia). In WP 2.4 (Clinical epidemiology and prospective registries for patients on standardised protocols) FORTH is enrolled in task 2.4.3 (**Implementation of a prospective WT clinical study in the ACGT system**). In this task USAAR is enrolled as an associate partner. The objective of this last task is to implement prospective clinical, imaging and biological data collection on patients treated according to the standard arms of the SIOP Wilms tumour 2001 clinical study using the ACGT platform. This will permit integration of complex data from the current clinical database with existing and prospectively acquired molecular biology data and imaging studies (DICOM data) to underpin identification and evaluation of biomarkers. The ultimate aim is to implement a Web-based integrated data collection tool for studies on Wilms tumour patients as a proof of principle for other prospective data collections and clinical trials in patient groups with a very favourable prognosis. The task description is as follow:

1. Pilot ObTiMA (Ontology based Trial Management Application) as an ACGT tool for data management on patients registered in the current SIOP WT 2001 trial and study.
2. Establish a DICOM-server and a system for international central imaging review.
3. Define a database for the logistics of bio-banking of Wilms tumour material throughout Europe to test the decentralised storage of Wilms tumour biomaterial.
4. Evaluate and enhance the ACGT Master Ontology for Paediatric Oncology
5. Use the ACGT workflow enactor to develop workflows for the analysis of molecular biological data derived from these Wilms tumours by the seamless integration of clinical, imaging data and web based data in a standardised way to define new risk factors for the stratification of Wilms tumour patients.

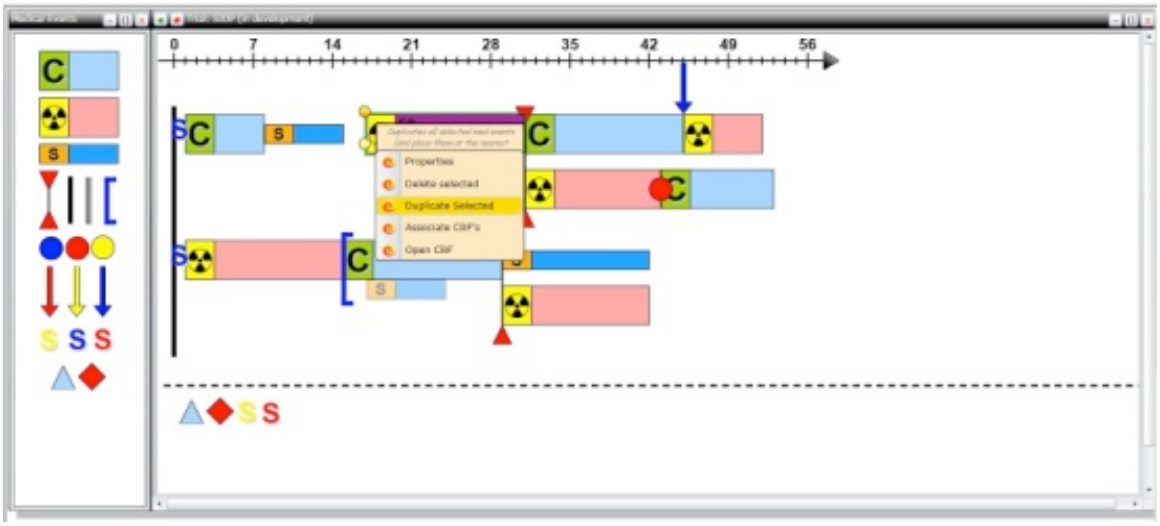
It has to be stressed that a prospective collection of clinical, imaging and postgenomic data needs a legal framework. This is especially true as these data are used by many different and sometimes multi-role endusers. Within this task the legal framework of ACGT will be tested based on contracts with hospitals, informed consents with patients, and IT tools for data security.

APPENDIX D: ObTiMA

ObTiMA is an **O**ntology-**b**ased clinical **Tri**al **M**anagement **A**pplication intended to support clinicians in both designing and conducting clinical trials[6,7]. The development of ObTiMA started in the ACGT project.

The design phase is facilitated by the Trial Builder in which all aspects of a clinical trial can be specified: a trial chairman can define the outline and metadata of a trial in a master protocol to describe, e.g., trial goals or administrative data. He can further setup treatment plans for guiding clinicians through individual patient treatment where events, e.g., surgery or chemotherapy, can be defined with all necessary information. Also, the particular treatment order can be freely setup on a timeline as well as treatment stratifications and randomizations to be applied for a patient. A Case Report Form (CRF) can be assigned to each treatment step to collect documentation data[8].

Figure D.1: Associating CRFs to a medical event and an example of how a treatment plan could look in TOB



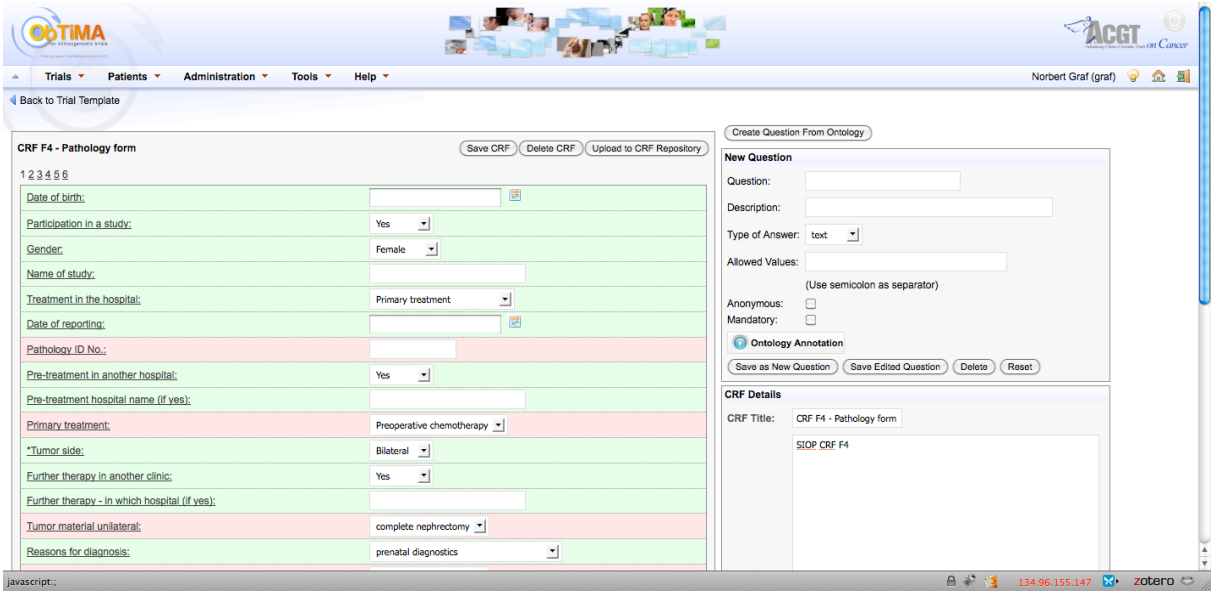
The ontology-based creation of CRFs in the Trial Builder is one of ObTiMA's major functionalities. A graphical user interface allows defining content, navigation, and layout of CRFs to capture all patient data during a trial, e.g., medical findings or diagnostic data. The

- [6] Weiler G, Brochhausen M, Graf N, Hoppe A, Schera F, Kiefer S: Ontology Based Data Management Systems for post-genomic clinical Trials within an European Grid Infrastructure for Cancer Research. Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, August 23-26, 2007, SuA11.4; Conf Proc IEEE Eng Med Biol Soc. 2007;1:6434-6437
- [7] Brochhausen M, Weiler G, Schera F, Rauch J, Graf N, Kiefer S: Ontology-based Trial Management System (ObTiMA). Nature Precedings http://dx.doi.org/10.1038/npre.2009.3753.1, 2009
- [8] Kuwahara M, Tanaka Y, Sjöbergh J, Graf N: Trial Outline builder. Flexible software tool for building, managing and analyzing clinical trials. HealthGrid conference 2010, submitted.
-

resulting descriptions are based on ACGT Master Ontology concepts for each CRF item along with metadata, e.g., data type and measurement unit, and used to setup the trial database.

The user interface makes the underlying aspects of the ontological metadata transparent to user and tries to overcome the gap between clinical practice and the actual logical representation of ontological concepts. (Even if natural language descriptions are given for concepts, those often do not mirror the needs of clinical perception of reality. Therefore an application-specific, simplified view of the ontology is given showing only its relevant portions in a clinician-friendly way.) When an item has been created based on a concept, its attributes are determined automatically, e.g., label, data type or answer possibilities but can be manually adopted.

Figure D.2: Creating a CRF in ObTiMA. Items with a green background are linked to the ontology whereas those with a red background are not linked to the ontology.



Since many trials collect similar data, it is possible to store components of or complete CRFs in a repository as templates. When setting-up a clinical trial, fitting CRF templates can either be directly reused or can be quickly created by plugging together existing CRF components. This in turn fosters the CRF standardization since CRFs can then readily be compared on the level single items (through ontological concepts) and also on the component level or in their entirety.

The second major functionality is the patient data management system supporting clinicians during a clinical trial. It is automatically set-up based on the items defined in the Trial Builder in the design phase. It guides the clinicians through the treatment of the individual patients according to the given treatment plans and provides an easy user interface to fill in the CRFs for a patient (again hiding the actual underlying ontology concepts from the clinician). When the PDMS (Patient Data Management System) is set-up, the trial database is automatically derived from the ontology-based CRF definitions. Thus, given appropriate rights are given, the database can then also be accessed by other trials or applications through using a semantic mediation service based on the ontology.

ObTiMA itself is composed of different modules. Besides the above described basic components a DICOM server and DICOM viewer, a SAE and SUSAR reporting tool and a consultation tool are integrated. These tools are optional to handle images used in clinical

trials, to simplify the SAE and SUSAR reporting according to GCP criteria. The consultation tool will store all consultations in a standardized way in the trial database. ObTiMA itself fulfils GCP criteria, including an Audit Trail. Data safety and security are guaranteed as pseudonymization of private data is done according roles and rights assigned to users of ObTiMA. It is of utmost importance that ObTiMA & TOB will be certified for the use within clinical trials.

APPENDIX E: Exploitation performance in the three exploitation frameworks

As part of the ACGT exploitation plan the consortium developed a set of criteria and metrics for managing exploitation-related work and activities in the three defined exploitation frameworks. This appendix presents these criteria in the *infrastructure* and *integrated services* frameworks and reports on the degree of achievement of each together with specific recommendations that can be drawn in light of the final results achieved.

E1: ACGT as an infrastructure

Criterion	Metric	Implication for ACGT	Achievements and Recommendations
Ease of development for the infrastructure	1. Underlying technologies used 2. Supporting documentation	I1: Ensure availability of documentation. The ACGT competition will act as a focal point for collecting, organizing and presenting this.	The ACGT experience has shown that certain technical selections made pose significant difficulties to 3rd parties wishing to develop for the ACGT infrastructure. Chapter 4 "Lessons learned" discusses this in more detail and recommends that a possible solution is the selection of middleware that is more flexible and open.
Number of end user resources	The more the better	I2: Resources already committed. The ACGT competition is expected to generate more.	The number of end user resources developed was sufficient to prove the concept of ACGT but not enough to allow integrated solutions to be offered. In accordance with our earlier finding on the necessity for well researched end user applications, we confirm both the criterion and its metric as valid.
Number of tools/resources that support the development of compliant services	The more the better	I3: Need to organise and present what exists in a more accessible manner. WP14 will be assuming a more active role in this.	This criterion has been achieved. Additional work on 'polishing' the applications in order to provide a smoother user experience certainly helps promote their usage.
Expected longevity of the infrastructure	1. Assurances for financial support of	I4: Visible and official partner commitments	This has been achieved indirectly via the success of a number of follow on

	infrastructure 2. Credibility of assurors 3. Track record	where possible	projects and efforts. Additional benefits can be gained by stressing the link of these efforts with ACGT so as to build on the momentum already generated.
Compliance of infrastructure with existing standards		I5: Document and present visibly in web, portal and other access points.	Where available, it is not easy to find this information. Proposal is to make it easily accessible from the top level of the ACGT site.
Infrastructure robustness	High score in robustness tests	I6: Test extensively and document problems and fixes where appropriate.	Achieved to a sufficient degree.
Number of existing users	Either a large number of users or a smaller number of high profile ones.	I7: The ACGT competition aims to address the first metric. The EORTC, NeoBIG and indeed University of Saarland are considered high profile end users.	A high number of end users has not been achieved. However high profile users have supported ACGT and will be exploited in follow up projects. Early awareness and recruitment of 3rd parties must be an essential goals of any such large infrastructure project.
Performance	Speed of applications	I8: It is early days and actually getting jobs done in the first place is more important than the speed at which they run.	Application performance is judged satisfactory in most cases, while in other it is exceptional.
Availability of system	1. Low downtime	I9: Monitoring system already in place. Possibly make this more prominent and open for public access.	Once stability of the various system components was achieved, system availability was also at a satisfactory level.
Supporting tools (e.g. usage and uptime monitoring, etc)	Number, findability and ease of use of these tools	I10: Document them extensively, make them easily searchable (via the Portal)	Most of the essential tools have been produced although their findability and ease of use are not as one would hope for. As such tools target 3rd party developers who are essential for the creation of the all important end-user applications, it is

			recommended to afford them the appropriate status when considering resource allocation.

E2: ACGT as an integrated environment of resources, tools and services for end users

Criterion	Metric	Implication for ACGT	Achievements and Recommendations
Nature of resources	• Resources must address a variety of actual, valuable tasks	I1: The consortium has selected 'usage scenarios' with the involvement of the end users. Need to use these with 'live data' and document as case studies.	While resources developed were useful on an individual basis, many of the original 'usage scenarios' were artificial, resulting in less than hoped-for uptake. Meeting true, present, valuable end user needs is absolutely critical.
Number of resources	• The more the better	I12: Most probably we need to create more resources. The ACGT competition aims to address this.	As above.
Supporting materials	• Documentation • On-line tutorials	I13: Create more supporting materials. (see WP14)	A significant amount of supporting materials have been produced. IN an era however where no-one reads the manual, finding appropriate forms of providing the necessary support is a real challenge.
Integration with legacy systems	• Integration should be transparent to end user	I14: No attempt to address this so far. Need to be aware of what is needed to achieve this integration. Probably need to create relevant documentation.	This was not required by any of the end users.
Integration with other ACGT services and resources	• Number of other ACGT resources with which each service can be combined to support more	I15: Basic integration achieved for the selected use-case scenarios. Need additional tests between all	Integration achieved within use case scenarios only resulting in a fragmented collection of services and resources. A powerful and user-friendly directory service

	complex workflows	available resources and services	together with flexible service interfaces is paramount.
Quality	• As measured by accepted metrics (application specific)	I16: Define metrics, create and organize tests as applicable, conduct tests and report findings	Work done in relation to some of the services (e.g. the literature mining modules) with favourable results.
Performance	• Application specific metrics. • Possibly availability of comparative information	I17: Define metrics, create and organize tests as applicable, conduct tests and report findings	As above.
Track record	• History of use	I18: Create a log of use and provide some basic access to it	Not done.
Real and perceived utility	• The resource is seen to offer true value to its end users	I19: Requires actual use by end users, creation of relevant evaluation questionnaires and analysis of results	End user assessments carried out during the final year of the project ad appropriate lessons documented in relevant deliverables (See WP 14)
Local support requirements	• Ideally these should be minimal	I20: Documentation of what is needed and list of solutions offered	Materials were produced in preparation for the ACGT competition.
Long term professional support	• Assurances of technical support by trusted assuror	I21: Visible and official partner commitments where possible	This is available on a case by case basis, usually by the developers themselves ad in the case of the commercial partners of ACGT.