

NEWSPAPER LETTER

ACGT

Advancing Clinical Genomic Trials on Cancer

EDITORIAL

As every project has an end, it is now time to say goodbye to ACGT as the project ends this summer, after four challenging and rewarding years. Building on the increasingly important ICT demand for innovative solutions for healthcare, ACGT has built a translational infrastructure for Clinical Trials for Cancer which will be resulting in applications that will increase the coherency of the entire system for building Clinical trials. Even if

the Islandic Vulcano has showed no mercy to the project's members scattered all around the world, ACGT demonstrates that willingness aligned with strong scientific orientation can make a project carry forward all of its intended results.

From Legal and ethical aspects to purely ICT solutions, ACGT legacy will be seen under several EC co financed projects such as ANCCA and CONTRACT, projects that will support the platform created by the ACGT project. We know that many partnerships developed during the project lifetime will continue and we hope that many of you will take the opportunity of this newsletter to get in touch with the project management to engage in further innovative discussions and partnerships.

This editorial to say goodbye to our readers and supporters is also the opportunity to thank the entire ACGT consortium and the individuals, user groups, organizations that have continuously supported the project during its lifetime. We wish you a very pleasant reading through the final ACGT Newsletter that we send you as a farewell and information on the ACGT project platform.

Truly yours
Samuel Keuchkerian

Newsletter Edition

Final Edition. Summer 2010

Editor-in-Chief

Samuel Keuchkerian
samuel.keuchkerian@health-grid.org

Editorial Board

Prof. Dr. Norbert Graf
Norbert.Graf@uniklinikum-saarland.de
Mr. Stelios Sfakianakis
ssfak@ics.forth.gr
Ms. Ana Lucia Da Costa
ana.dacosta@healthgrid.org

Contact ACGT

Web: www.eu-acgt.org
Scientific Director:
Dr. Manolis Tsiknakis
tsiknaki@ics.forth.gr
Administrative Manager:
Mrs. Jessica Michel
jessica.michel@ercim.eu

Newsletter Design

HealthGrid



European Commission
Information Society and Media



Editorial	1
Clinical Trials News	2
Products and Services	4
Grid News	5
Feature Article	6
Community View	8
Events	10
Legal and ethical	10
Life in ACGT	12
ACGT People	14
SPECIAL PROJECT CLOSING	15

Clinical Trials News

Latest developments in the world of clinical trials in cancer

The elaboration of guidelines and recommendations for integrating clinical data sources into the ACGT platform



An important challenge in carrying out post-genomic bio-medical research is to efficiently manage and retrieve all relevant data from many heterogeneous sources. A post-genomic clinical trial involves the collection, storage and management of a wide variety of data, including: clinical data collected on Case Report Forms (e.g. symptoms, histology, administered treatment, treatment response), imaging data, genomic data, pathology data and other lab data. Next to that, access to many external sources of data and knowledge is required. These store information about gene and protein sequences, pathways, genomic variation, microarray experiments, medical literature, etc. Seamless access to all these data repositories would greatly facilitate research.

In order to provide seamless data access, syntactic and se-

mantic integration needs to take place. Syntactic data integration handles differences in formats and mechanisms of data access, the fact that information can be represented in different ways, using different terms and identifiers.

To achieve syntactic integration, the data access services first need to provide a uniform data access interface. This includes uniformity of transport protocol, message syntax, query language, and data format. Through the ACGT syntactic access services data can be queried using SPARQL, thus hiding the different query mechanisms provided by the underlying databases.

In ACGT we have implemented syntactic data access services to access relational databases, DICOM image repositories and BASE microar-

ray databases. Relational databases can also be used to make data available that may not yet be stored in a relational database, but that can be mapped to the relational data model. This holds for data collected in files of various formats, such as Excel files, plain text files, XML files, etc.

Next, the data access services export the structure of the database using a common data model, together with possible query limitations of the data source. An RDF Schema of the data resources is exported on demand. Clients use this information for constructing queries, e.g. the semantic mapping editor uses this schema to provide the mapping to the ACGT Master Ontology. Finally, the data access services enforce the data source access policy, and audit access to data sources.

The main steps to integrate new sources with trial-specific patient data into the ACGT platform are the same for all data sources, irrespective of the type of data they store.

1. Export the data from the data source.
2. Anonymise and pseudonymise the data.
3. Determine who should be allowed access, and if need be, create the appropriate contracts and have these signed by all parties.
4. Set-up a database and import the anonymised data.
5. Set-up secure access to the database.
6. Create a data access service for the database.
7. Configure the GAS so that authorized users can access the data by way of the data access services.
8. Create the required semantic mapping so that the database can be queried using the ACGT Master Ontology.

The integration of new data sources is not only a technical task. Successfully making new trial data available also involves clinical and legal partners in ACGT.

We have also provided users with the ability to dynamically deploy new relational data sources into the platform. Users can integrate new data sources from the ACGT portal and subsequently query these from the workflow enactor.

Dynamic creation of relational data access services is particularly useful because a lot of different types of data can be made available this way, including data from Excel spreadsheets and text files in CSV format. Integration of new data sources requires creating a mapping from SPARQL to SQL, which requires good familiarity with the schema of the database and the data contained in it. In contrast, there is no content-specific knowledge required to integrate new DICOM and microarray databases. The data they store is much more

specific and new databases appear less frequently (generation of new data in both cases requires expensive equipment, is time-consuming and always involves patients or tissue samples), which means that static deployment of these data access services is in practice sufficient.

*Anca Bucur
Philips*



Products and Services

News on the latest products or services in our area of interest

From Prototype to Production – Changes to the ObTiMA development to make it ready for clinical scenarios

When the development of ObTiMA was launched early within the ACGT project, it was first intended to become a showcase application for the various innovative technologies created within that project. The main focus at that time was to highlight the novel possibilities for designing and managing the various parts of clinical trials based on ontological concepts and descriptions. But very soon it became obvious that ObTiMA has a much greater potential than just being a research prototype. And therefore the decision was taken to develop the system further into the direction of fulfilling the strict software requirements found in clinical settings. But to create such a production-ready application the previous research-focused development process had to be shifted towards a commercial software development methodology. Also different criteria gained focus that are of lesser importance in research but essential in everyday use, such as usability (since in clinical environments, easy-and-quick-to-use software is of utmost importance). In the following we describe some of the measures taken over the last year in light of the just said:

As first step, the previous development process and the responsibilities of each participant were analyzed. For this, the original goals for ObTiMA (as described in the DOW and the deliverables) were “dissected” and put in relation to the current state of the system: Which goals were reached already, which ones should be realized, and which ones should be cut towards the goal of a stable system. The continued development should not attempt to include all imaginable, “cool” functionalities but rather concentrate on strengthening the major, critical functionalities retaining a sensible and realistic timeframe. In this context, a continuous communication between all developers was started to include regular face-to-face meetings and conference calls to discuss such planning together and in turn improve the team

networking. The planning is also fostered by a new centralized Wiki software (Atlassian Confluence) where all project documentation now takes places (and thus no more searching in old e-mails for some specification PDF).

One larger issue previously was the lacking of common guidelines on the different levels of the system development. Hence each developer was using his own coding-style together with his own preferred programming packages (such as for XML processing). To change this we refactored the existing code to use only common packages and teach the developers how to apply a coherent code style with e.g. intelligible file or method names and simplified code structure. This was combined with the introduction of a multi-level testing framework: For this framework, first, a contiguous set of sensible system data was developed that allows creating test cases with always reproducible results. This data, in turn, is applied by an automated testing environment that checks both the correct working of the underlying code as well as of the user interface. If those tests reveal any bugs in ObTiMA then those can be directly reported within the new bug-tracking software (Atlassian JIRA) which allows tracing the state of all reported bugs (as well as feature requests).

*Holger Stenzhor,
USAAR*



Grid News

Latest developments in the world of computer grid research

Proper launch of the EGI - A challenge for the European Grid Community

While grids are commonly understood as distributed computing infrastructures, they are primary about the collaboration. It is therefore not surprising that the first international pusher for the large scale grid infrastructure was a community behind the high energy physics scientific experiments. A development that initially started as a support activity of one scientific community became soon a movement that attracted scientists from many other disciplines and a recognition at the highest national political levels.

The initial efforts culminated in a series of the EGEE projects, run between 2004 and 2010 and supported by the European Commission within the 6th and 7th Framework Programmes. While successful, these projects also demonstrated the shortcomings of an infrastructure based on standard projects and taken care of by very heterogeneous group of supporting institutions, ranging from university departments to proper infrastructure building bodies.

The community, with the support from the commission, started in 2007 a design study project EGI_DS—European Grid Initiative Design Study—to properly define the conditions and requirements for a large scale trans-European grid infrastructure. The project, coordinated by CESNET and led by the author of this article, produced a series of documents, with the EGI Blueprint and accompanying

deliverables on functional requirements setting the framework for the future grid infrastructure in Europe. The proposal prepared by the EGI_DS was inspired by the successful history of trans-national computer networks and their organizational model, based on a combination of a national and international coordination. The basic building blocks for the future European Grid Infrastructure were therefore defined as National Grid Initiatives (NGI), organizations that are responsible for the national coordination of grid infrastructure and that are also representing individual countries at the international level. The NGIs are complemented by the EGI.eu, an organization founded as part of the EGI_DS activities, with headquarters in Amsterdam in The Netherlands. EGI.eu together with the NGIs form a basic foundation of the European grid collaboration and are responsible for the coordinated operation of the European Grid Infrastructure. The governance is guaranteed through the EGI Council, where each participating NGI, as well as international body like CERN, does have its representative.

The long term sustainability of this model should be provided through national funding of NGIs that understand benefits of close international collaboration and coordination of activities. The first step towards this model, the so called Transition towards the EGI, is financially supported by

European Commission through its moderate co-funding (at the level of one third of total expenses) for a four year EGI InSPIRE project. The project started on 1st May 2010 and its success or failure will determine not only the validity of the EGI model, but the fate of the grid EGI collaboration itself. With NGIs and not individual research institutes its primary partners, the infrastructure managed by the EGI InSPIRE is much more application neutral than its EGEE predecessor. While the project itself includes a very basic funding for the heavy user communities like the HEP, Bioinformatics or Computational Chemistry, the major challenge lies in providing an infrastructure that will be attractive for scientists irrespective of their organizational and scientific affiliation. A close collaboration with user communities—most notably those behind large ESFRI projects—is essential, but EGI must not abandon also smaller and more dispersed scientific communities. This task is even more challenging as the major projects targeting these communities were not accepted by the Commission and EGI InSPIRE must look for new partners and new ways of collaboration with user communities. One way to do this is through emphasizing the roots—grids as a collaboration platform, connecting scientists across Europe. Properly managed infrastructure open to any scientist pave the way towards trans-European grid infrastructure that will be widely accepted as the basic eInfrastructure of the future, with no fear for its sustainability. Let's support the EGI InSPIRE project to properly respond to this challenge.

*Ludek Matyska,
Institute of Computer Science
Masaryk University, Czech Republic*

Feature Article

ACGT links to the new EU funded projects: ENCCA and CONTRACT

The necessity of an environment that deals with the recent advances in high throughput genomics and post-genomics methods and technologies in molecular biology, which have resulted in an explosion of information and knowledge about cancer and its treatment, is more than evident today. As a result of this explosion in information and knowledge, our ability to characterize and understand the various forms of cancer is growing exponentially. Information arising from post-genomics research and combined genetic and clinical trials on one hand, and advances from high-performance computing and informatics on the other is rapidly providing the medical and scientific community with an enormous opportunity to improve prognosis of patients with cancer by individualizing treatment. Multi-level data collection within clinico-genomic trials and interdisciplinary analysis by clinicians, molecular biologists and others involved in life science is mandatory to further improve the outcome of cancer patients. It is essential to merge the research results of biomolecular findings, imaging studies and clinical data of patients and to enable users to easily join, analyze and share even great amounts of data.

The vision of the European Network for Cancer Research in Children and Adolescents (ENCCA) funded by the European Commission as a Network of Excellence (NoE) is to establish a durable, integrated clinical and translational research infrastructure for Europe that will define and implement its research strategy and will facilitate the necessary investigator-driven clinical trials to introduce the new generation of biologically targeted drugs into standard of care for children and adolescents with cancer. As the goal of ACGT is to develop an open-source and open access IT infrastructure that provides the biomedical research community with the tools needed to integrate complex and heterogeneous data from clinical information systems, from molecular biology (post-genomic data), from imaging studies (DICOM data), and from the Web in a standardized way FORTH as the scientific coordinator of ACGT was asked to participate in the NoE as a funded member of the consortium. USAAR as a partner in ACGT is a non-funded member of ENCCA. The ENCCA NoE brings together **internationally recognised Paediatric Oncology Academic Institutions** and many associated organisations that aim to integrate the existing **Clinical Trials Groups in Paediatric and Adolescent Oncology** towards a **European Virtual Institute** to reduce knowledge fragmentation and enhance their communication, collaboration and management of effective clinical research in Europe. The objective is to **structure knowledge-sharing** through the integration of the whole chain of stakeholders

(epidemiologists, imaging developers, biologists, clinicians, drug developers, industrials, parent and patient groups, ethical, IT and regulatory authorities). This will **support the development of specific therapeutic strategies relevant for tumours** in children and adolescents and should accelerate the development of targeted compounds for cancers in this age group. The contribution of ACGT in this NoE will be in helping to establish an integrated ICT infrastructure for Paediatric Oncology in Europe. Specifically FORTH will lead a task dealing



with the analysis of existing ICT systems and solutions for patient administration and clinical data management and together with USAAR FORTH is enrolled 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 this task ObTiMA will be piloted for data management on patients registered in the current SIOP WT 2001 trial and study. A database for the logistics of bio-banking of Wilms tumour material throughout Europe will be defined to test the decentralised storage of Wilms tumour biomaterial. The ACGT workflow enactor will be used 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. During the life time of the project the ACGT Master Ontology will be evaluated and enhanced for Paediatric Oncology. In addition a DICOM-server and a system for international central imaging review will be build. 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 end-users. Within the NoE the legal framework of ACGT will be tested based on contracts with hospitals, informed consents with patients, and IT tools for data security.

CONTRACT's acronym stands for "Consent in a Trial and Care Environment" and consent will be the main focus of the project. The Consortium - involving five partners from universities, research institutions and SMEs of which three belong to the ACGT project, namely, LUH (project coordinator), USAAR and Custodix, will focus on analyzing how the legal concepts of informed consent in the European Data Protection Directive and in the Clinical Trials Directive differ from each other and from the different regulations of informed consent for care in chosen Member States of European Union. The project will analyse the consent issues in vulnerable patient groups, with the paediatric population in the central point as the requirements are of the most demanding complexity there.

The outcome should be an assessment how far those different understandings of informed consent have an impact on the success of translational research. The concept will be analyzed from a legal, ethical, IT-related and clinical point of view. The Consortium will use the outcomes of their own research in the ACGT project (See for example <http://eu-acgt.org/documents/public-deliverables.html#D101>, <http://eu-acgt.org/documents/public-deliverables.html#D102>).

CONTRACT will support translational research projects – both ongoing and upcoming. It will develop a multidisciplinary approach in delivering facts and figures on different approaches to informed consent both in European projects and in European Member States. CONTRACT will analyze the IT-related representation of these different understandings and the outcome of these differences in the daily clinical and/or research routine. To achieve this the consortium will circulate a questionnaire directed to the partner projects which are, or were working with vulnerable patients. In those questionnaires the questions about legal and ethical, security and clinical practices will find their place.

CONTRACT will in further steps advise translational research projects in all issues of informed consent and will deliver concrete policy recommendations as to how the European Union could jointly protect patient's rights and support translational research by a better structured approach towards consent issues.

A helpdesk will be established that it will offer advice to projects on all possible issues of informed consent. The helpdesk will offer a data-protection-framework ready to run and will provide a help forum for legal, ethical, IT-related and clinical questions related to informed consent and data protection in translational research.

Finally the project results will be disseminated at two stakeholder workshops to be held in month 12 and 24 of the project. These will provide an ideal forum for European policy makers and researchers to exchange views on the project's results and possible options for future policies in this area.

Nikolaus FORGO, LUH
Manolis TSIKNAKIS, FORTH

Community View

Invited contributions from non-ACGT members of the wider research community

Enhancing Access to Cancer Clinical Trials

The Education Network to Advance Cancer Clinical Trials (ENACCT) is a US-based nonprofit organization, founded 2004 with support from the Lance Armstrong Foundation. ENACCT's mission is to improve access to cancer clinical trials through education and collaboration with communities, health care providers, and researchers. ENACCT helps patients, communities, and health care providers better understand the importance of cancer clinical trials—while also helping researchers communicate more effectively about these trials.

In this article, I am pleased to share ENACCT's perspective on what it takes to enhance access to cancer clinical trials.

1. Work to Create Community Literacy About Clinical Research:

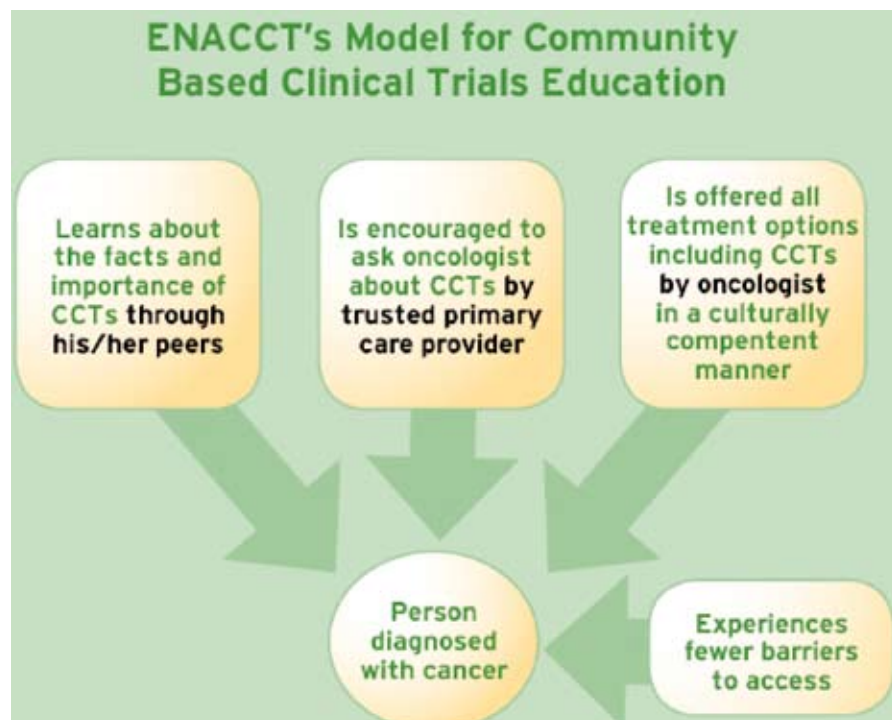
We need to consider that the public knows very little about clinical trials, and what it knows is extremely negative. At ENACCT, we believe that the optimal “teachable moment” may not be at the time of diagnosis, but may be at other times, such as churches, health fairs or community events, to help everyone understand that when a mother, sister or neighbor is diagnosed with cancer, she should ask her doctor, “Is there a clinical trial that’s right for me?” Considering the patient in the center of our educational programs as illustrated in the diagram below, ENACCT has developed a set of unique training programs to help

- a) Community **leaders reach their peers** with simple messages focusing on the importance of clinical trials as a part of quality care.
- b) Primary **care providers talk with their patients** prior to referral to oncology about the importance of clinical trials as a quality treatment option
- c) Clinical **trial staff** to better conduct outreach, recruitment and retention for their trials.

While most of our programs are delivered live, we have a number of free on line courses that include Continuing Education Units for physicians, nurses and social workers. See <http://www.enacct.org/our-programs/your-role-cancer-clinical-trials> for more information.

2. Consider how clinical trial messages are portrayed in publications

Many publications discuss treatment as somehow “separate and apart from” clinical trials. However, for many people with cancer, clinical trials can be a viable treatment option. Why do we always save discussion on this issue to the moment of “last resort?” Take a look at how your own organization discusses clinical trials. We need to make sure that language we use helps give confidence to all patients to ask their physician, at all stages of treatment, “is there a clinical trial for me?”



Since founding the US-based ENACCT in 2004 with funding from the Lance Armstrong Foundation, Margo Michaels has been the organization's Executive Director, whose mission is to improve access to cancer clinical trials through education and collaboration with communities, health care providers, and researchers.

3. Consider how patient advocates and community representatives can become more involved in clinical trials design and implementation

Less than three percent of all adult cancer patients participate in clinical trials. According to recent research by Dr. David Dilts from Vanderbilt University, between 50 and 60 percent of all the trials entered in US based cancer centers accrue fewer than five patients, and approximately one quarter accrues no patients! To address these concerns, a number of reports have called for greater inclusion of public representatives in clinical research design and implementation.

In 2007, we spearheaded a new effort called Communities as Partners in Cancer Clinical Trials: Changing Research, Practice and Policy, along with Community-Campus Partnerships for Health (CCPH), with core funding from the US government. Our final report highlights 58 recommendations for community engagement in the cancer clinical research process – from trial design to implementation to dissemination of results -- in order to address low accrual rates in cancer clinical trials-- and ultimately improve research outcomes. It is available at www.enacct.org. Our report is the first to detail why and how the cancer clinical trial process can involve communities affected by cancer– from trial design to implementation to dissemination of results - with a focus on community engagement strategies. We encourage you to “lift” the ideas that work for you and your country.

*Margo Michaels,
Executive Director ENACCT*



In ENACCT communities, we hope to address patients' lack of awareness about clinical research through reaching the groups that he or she is likely to encounter throughout the diagnostic process. The newly diagnosed patient will a) first hear about clinical trials through his/her peers, whether in the neighborhood, at the beauty shop, at church, or through an advocacy group; b) will then learn about the option of receiving treatment through a clinical trial by a trusted primary health care provider; c) will be better

able to initiate inquiries about clinical trials and d) will be offered the opportunity to join a clinical trial by the oncologist, if he/she is eligible.



Events

Information on upcoming events of interest

VPH 2010

30th September to 1st October, 2010,
Brussels, Belgium
<http://www.vph-noe.eu/vph2010>



The Virtual Physiological Human Network of Excellence (VPH NoE) will hold the first of a series of VPH Conferences (VPH2010) on 30th September to 1st October, 2010. The first meeting will be held in Brussels, Belgium, and is supported

by the European Commission ICT for Health / DG Information Society and Media. The Virtual Physiological Human Network of Excellence is an umbrella project representing the Virtual Physiological Human Initiative set up by the European Commission with a budget of ~ 350 million Euros for the Framework 7 Program. The VPH NoE is also responsible for producing a 'VPH Vision and Strategy Document' to inform the next calls for the VPH field for FP7, FP8 and for European large infrastructural actions.

The meeting on 30th September - 1st October will be dedicated to the VPH Initiative, bringing together key representatives from VPH groups, Industry and Clinics. The first meeting is designed to be a relatively small meeting up to a maximum of 200 participants with high profile key speakers and will contain the very best of VPH research. Parallel sessions will be held on both days with proposed topics in four main themes; Organ Systems, Modelling Scales, Research Methodology, Application Domains and Data Access and Information Assurance. We anticipate increasing the size and scope of the VPH conference in subsequent years.

For further details, please visit:
<http://www.vph-noe.eu/vph2010>

Medinfo 2010

12 - 15 of September 2010
Cape Town, South Africa
<http://www.medinfo2010.org>

Cape Town, South Africa will host the 13th International Congress on Medical Informatics from the 12 - 15 of September 2010. The Medinfo conference is the premier triennial international meeting for the medical informatics community. It brings together world leaders in this field to share knowledge and experiences. Medinfo 2010 is a unique opportunity to meet these leaders and to hear of, and contribute to, advances in biomedical and health informatics. The Medinfo conference is the official conference of the International Medical Informatics Association (IMIA), a bridging organisation aiming at supporting stimulating high-quality translational communication, research, education, and practice in biomedical and health informatics. Medinfo 2010 is organised and coordinated by the South African Health Informatics Association. A number of governmental organisations, associations and businesses have agreed to be sponsors. The conference theme selected for Medinfo 2010 is a topic that most countries are currently addressing: Partnerships for effective eHealth solutions - Innovative collaborations promote solutions to health challenges.

For further details, please visit:
<http://www.medinfo2010.org>

Legal and ethical

The latest thinking on legal, ethical and data security issues surrounding clinical trials

ACGT: Risky Business?

Safeguarding the privacy of patients is an important task for anyone dealing with medical data as harming it can do tremendous damage to the individual involved (note that privacy breaches are irreparable as an information leak cannot be undone). With respect to the research itself, such incidents are likely to lead to patients withdrawing from participation

(damaged trust) and could very well lead to prosecution of those responsible for the breach.

The ACGT project has contributed considerable effort to solving the data protection-related issues which accompany the creation of a transnational biomedical research infrastructure. The project aimed to provide researchers with a convenient and easily implementable solution to deal with legal and regulatory compliance regar-

ding data privacy. The solution has been defined in the form of a framework consisting of a combination of technical, organisational and legal measures.

This ACGT Data Protection Framework (DPF) builds upon the concept of "context of anonymity", the establishment of a Data Protection Authority (DPA) and on the integration of a Trusted Third Party. In short, the main goal of the DPF is to ensure that within the ACGT environment, data can be considered "de-facto" anonymous and as such it frees researchers from the tedious administrative

tasks connected to data protection compliance for each individual research initiative, as long as they stick to the global ACGT rule set.

A key factor in determining the success of this approach is the question to what extent the DPF offers guarantees with respect to regulatory compliance. To answer this question and assess the strengths and weaknesses of the DPF, a risk analysis concerning data security and data protection was performed. The four main threats that were examined are:

- Non-compliance
At all levels, be it technical, procedural or contractual.
- Infrastructure and technology-related security risks
- Contextual anonymity
Specific risk associated with the establishment of a controlled context in which data can be considered de-facto anonymous.
- Long-run sustainability
Risks associated with maintaining the research infrastructure on the long term.

Within ACGT a set of interdependent base components forms the foundation of the entire infrastructure. The security related components are largely based on a set of proven concepts, technologies and implementations. Risks associated with purely technological aspects are therefore limited.

However, technical security measures leave a number of gaps open due to architecture and design limitations or simply because it is technically impossible to prevent certain actions. For example: a system adminis-

tration can forbear to disable accounts of people leaving an organisation (not following employee removal policies); those people can still access data, although they are no longer authorised.

To cover these gaps, the DPF resorts to procedures, contracts, End User License Agreements (EULA) and other (legal) agreements. Non-compliance with these procedures (related to maintenance and management of operations) and contracts (legal requirements imposed by ACGT) pose an increased risk. Irrespective of the cause of this non-compliance (ignorance, sloppiness or malicious intent), the consequences can be devastating as a technical security architecture can only function properly if the legitimate users adhere to the policies.

This mainly “human factor” is the biggest risk to data privacy. The sheer size of the project (cf. for one, the geographical distribution of people involved) is an additional complicating factor to controlling compliance. Controlling this risk includes preventive measures such as

training and avoiding procedural complexity of procedures (a goal not easily reached) and corrective measures such as financial penalties (the ACGT agreements for participation foresee such penalties).

On the plus side, the risk analysis shows that most identified threats are associated with the leaking of de-identified data (thanks to the DPF). In general these will have a limited impact as it is highly unlikely that data will end up with people capable of re-identification (especially not with accidental breaches).

And finally, again the need for a separate governing body such as the Center for Data Protection (CDP) in international collaborative environments can be concluded, because in a long-time operational infrastructure, there is a clear need for coordinated technical and procedural auditing.

*Brecht Claerhout,
Custodix*



Life in ACGT



Project coordinator's Farewell

ACGT (Advancing Clinico-Genomic Trials on cancer: Open Grid Services for improving Medical Knowledge Discovery) is an Integrated Project (IP) funded in the 6th Framework Program of the European Commission under the Action Line "Integrated biomedical information for better health".

Its high level objective has been the development of methods and systems for improved medical knowledge discovery and understanding through integration of biomedical information (e.g. using modeling, visualization, data mining and grid technologies). Biomedical data and information that have been considered include clinical information relating to tissues, organs or personal health-related information, but also information at the level of molecules and cells, as acquired from genomics and proteomics research.

The project vision has been rooted in the realization that information arising from post-genomics research and genetic and clinical trials is rapidly providing the medical and scientific community with new insights, answers and capabilities when combined with advances in high-performance computing and informatics.

The over-arching objective of the ACGT project has been the provision of a unified technological infrastructure which facilitates the seamless and secure access and analysis of multi-level clinico-genomic data enriched with high-performing knowledge discovery operations and services.

During the course of the four and one-half years of its life, the project has defined a detailed architectural blueprint and has developed, tested and validated a range of technologies, such as:

- new, domain-specific ontologies, built on established theoretical foundations and taking into account current initiatives, existing standard data representation models, and reference ontologies;
- innovative and powerful data exploitation tools, for example multi-scale modelling and simulation, considering and integrating from the molecular to the systems biology level, and from the organ to the living organism level;
- standards for exposing the properties of local sources in a federated environment;
- a biomedical grid infrastructure offering seamless mediation services for sharing data and data-processing methods and tools;
- advanced security tools including anonymisation and pseudonymisation of personal data according to European legal and ethical regulations;
- a Master Ontology on Cancer and use standard clinical and genomic ontologies and metadata for the semantic integration of heterogeneous databases;
- an ontology based Trial Builder for helping to easily set up new clinico-genomic trials, to collect clinical, research and administrative data, and to put researchers in the position to perform cross trial analysis; and
- data and literature mining services in order to support and improve complex knowledge discovery processes.

The technological infrastructure has been validated in a concrete setting of advanced clinical trials on cancer. Pilot trials have been selected based on the presence of clear research objectives, raising the need to integrate data at all levels of the human being. The project has targeted two major cancer diseases: breast cancer (BRCA) and paediatric neuroblastoma (PN).

In achieving these ambitious objectives, the project has brought together internationally recognised leaders in their respective fields, with the aim to deliver to the cancer research community an integrated Clinico-Genomic ICT environment enabled by a powerful grid infrastructure.

The ACGT consortium has pursued a long term vision to develop open-source, semantic and grid-based technologies in support of post genomic clinical trials in cancer research. Viewing the publications achieved over the life of ACGT we can say that lessons have been learned and good results were obtained.

Although the ACGT project is officially ending soon, the excellent research partnerships developed during the project will continue. The vision of becoming a pan-European voluntary network connecting individuals and institutions while enabling the sharing of data and tools and thus creating a European-wide web of cancer clinical research has been well advanced. The project has developed long lasting partnerships with some of the major stakeholders in the European Cancer Research arena, including ECCO (European Cancer Organisation), BIG (Breast International Group), SIOPE Europe (The European Society for Paediatric Oncology) and the European Clinical Research Infrastructures Network (ECRIN). Building upon the technologies, procedures and knowledge generated by the project, several ACGT partners – jointly with such important end user groups – are about to enter the second phase of implementation. This has been made possible through additional funding from the EU research programmes (both HEALTH and ICT).

In parallel to these developments, the primary instruments for ACGT's collaborative exploitation are well developed. The Center for Data Protection (CDP) has been established and is already actively engaged in service provision. Also the STaRC Initiative has grown into maturity. STaRC is intended to be a 'Study, Trial and Research Centre' that will exploit clinically relevant aspects of ACGT.

The concept behind STaRC has received significant recognition and support from patient organisations and patient support groups as well as from some regional governments. The activities for its official initiation are almost complete.

As the people that shared the coordination and management responsibilities of ACGT, we look back over the project lifetime remembering the many challenges that we faced and the many exchanges on very critical issues. Our conclusion is that the biggest strength and benefit of the ACGT project has been its people.

As you move on to new projects and new endeavors, we hope you will take with you fond memories of the achievements that you accomplished and the people that you worked with to accomplish them in the context of ACGT. We would like to thank the consortium members heartily for their hard work and enduring dedication to the project.

Many of us will meet one last time for the ACGT Final Review, to take place in Heraklion September 22-23. Until then, we wish you safe and happy summer holidays!

Manolis Tsiknakis, FORTH, Scientific and Technical Coordinator
Jessica Michel Assoumou, ERCIM, Administrative and Financial Coordinator

ACGT PEOPLE



Yannick Legré

holds an Engineer Degree in Telecommunication and Networks. He is one of the co-founders and current president of the International HealthGrid Association. He works in the domain of Grid technologies applied to healthcare and biomedical research since August 2000 and he has been successfully involved in more than 15 projects. He is regularly acting as an expert for the European Commission. Since June 2008, he is Director of International Relationships for the maatG company (www.maatg.com)



Andreas Persidis

is the CEO of Biovista, the ACGT partner responsible for the literature mining, ontology browsing and related ACGT applications. Andreas' research interests focus on knowledge management and the development of tools for systematic knowledge discovery in the life-science domain. With a long career in artificial intelligence, ontology development and text mining but also significant business experience, he serves as an expert reviewer and advisor to national organizations, investment banks and the EC itself. Andreas has brought this experience to bear within ACGT, pushing the consortium to always take into account the true needs of the stakeholders and to develop solutions that make sense in actual environments rather than purely research contexts. Not being any longer able to program to save his life, Andreas instead works on spreading the word for the better exchange and use of data to address the complex issue of better health care. He believes that mining the information in patient records is a step in the right direction. Andreas shares his time between Europe and the US and what time remains free, he enjoys with his 5 year old daughter Leda who has yet to see tangible benefits for her from his ACGT work.

SPECIAL PROJECT CLOSING

The TOP trial for the final ACGT Demonstration

The ultimate objective of the ACGT project is the provision of a unified technological infrastructure which will facilitate the seamless and secure access and analysis of multi-level clinical and genomic data enriched with high-performing knowledge discovery operations and services, in the concrete setting of **clinical trials on Cancer**. Pilot trials have been selected based on the presence of clear research objectives, raising the need to integrate data at all levels of the human being. This integrative view underlies the development of clinico-genomic models, showing that the combination of biomarkers and clinical factors are most relevant in terms of statistical fit and also, more practically, in terms of cross-validation predictive accuracy.

The **TOP trial**, a trial which aims at identifying molecular markers that predict response/resistance to one of the most commonly administered chemotherapies in breast cancer, is one of those pilot trials for ACGT and has been chosen for the final demonstration of ACGT.

To date very little progress has been achieved in the field of biomarkers predictive of chemotherapy benefit in breast cancer. Consequently, the vast majority of patients considered to be at moderate or high risk of relapse are treated with the cytotoxic agents viewed as the

most active “on average”, namely anthracyclines and taxanes chemotherapy agents. These drugs have significant side effects, the most worrisome of which are secondary leukemias and irreversible congestive heart failure for anthracyclines, and slowly reversible neurotoxicity for the taxanes.

In the pre-operative international TOP trial presented here, we focus on identifying molecular markers that predict response/resistance to anthracyclines, one of the most commonly administered chemotherapies in breast cancer.



Interestingly, the TOP trial avoided any possible confusion by including only women being treated with single-agent epirubicin. We also limited the patient population to women whose tumours did not over-express the estrogen receptor (ER). This is because, for these women, chemotherapy can stop the ovaries from working, offering an additional benefit that may distort the results.

The aim of the trial was to carry out the first prospective evaluation of the predictive value of topoisomerase II alpha (TOP2A) gene aberrations and expression. TOP2A is a key enzyme in DNA replication, one of the molecular targets of anthracyclines, and it is amplified in 24% to 54% of HER2-amplified tumors. Although TOP2A is considered by some investigators to be a promising marker for predicting the activity of anthracycline-based regimens, inconsistent results have been reported about TOP2A amplification/expression and response to anthracyclines.

The study protocol also included exploratory analysis to

identify gene expression signatures that correlate with pCR. We therefore aimed to develop a gene expression signature to identify those patients who would not benefit from anthracyclines and could therefore be spared the non-negligible risks of this type of chemotherapy.

For the final demonstration of ACGT, it has been decided to use the TOP trial to illustrate several, but not all, procedures and tools set up during the ACGT project. This demonstration has been designed as a step by step process simulating how the TOP trial data is introduced and analyzed in the ACGT infrastructure to identify the targeted biomarkers.

In practice, we will first describe and illustrate the **legal framework** which is necessary to take into account the needs of modern scientific genetic research and the needs of the patients regarding data protection and privacy. This legal framework specifically involves the informed consent procedure which was carried out in the TOP trial, the collaborative contracts which had to be developed and agreed by

the data providers and data users, and the anonymization of the patient data.

Second, we will illustrate the importance of the **ACGT Master Ontology** for the semantic integration of heterogeneous data (clinical, imaging, genomic, proteomic, ...).

Third, we will demonstrate how tools developed within ACGT can facilitate the **identification of predictive markers of response/resistance** for anthracyclines chemotherapy using microarray-based gene expression profiling as well as genotyping technology.

Fourth, we will report the progresses and advances made by the **in silico oncology** working group. This group evaluates the reliability of in silico modelling as a tool for assessing alternative cancer treatment strategies; especially in the case of combining and utilizing mixed clinical, imaging and genomic/genetic information and data.

Since, until now, we were lacking **patients' views on and experience of involvement in clinico-genomic research**, Grid structures, and Europe-wide data flows, we will finish by reporting the results of the large empirical survey, carried out within ACGT, on perspectives and needs of persons who did consent to take part in tissue-based cancer research in several European settings.

*Christine Desmedt,
IJB*



JOIN ACGT

Membership in ACGT is open to all. Here are some benefits you enjoy as an ACGT member:

- Access to all member resources
- Support in solving problems in the areas of interest of ACGT
- Direct contact with ACGT experts in a variety of fields including clinical trials, cancer research, advanced software development, Grid implementations, legal, ethical and data security issues and much more
- Ability to contribute to the ACGT infrastructure and receive support for it.

**Join us at:
www.eu-acgt.org**

SUBSCRIBE!

Subscribe today and be among the first to learn of all the latest developments in ACGT and post-genomic clinical trials research on Cancer

3 EASY STEPS

1. Visit us at www.eu-acgt.org
2. Type in your name and email
3. Press the Subscribe button