HealthGrid Applications and Technologies Meet Science Gateways for Life Sciences
S. Gesing et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-054-3-3

HealthGrid 2012

Tristan GLATARD,[a,1] Silvia D. OLABARRIAGA[b] and Tony SOLOMONIDES[c]
[a]*Creatis, Université de Lyon, CNRS, Inserm, 69621 Villeurbanne, France*
[b]*Academic Medical Centre,University of Amsterdam, The Netherlands*
[c]*CCRI, NorthShore University HealthSystem, Evanston, IL 60201, US*

With HealthGrid 2012 (http://amsterdam2012.healthgrid.org) we celebrate the tenth meeting of this open forum for the integration of grid technologies, e-science and e-health methods and their applications in the biomedical and healthcare domains. The principal objective of the HealthGrid conference, as it was of our community, the HealthGrid Association, remains the exchange and debate of ideas, technologies, solutions and requirements that interest the grid and the life sciences communities and are likely to promote the integration of grids into biomedical research and healthcare in the broadest sense. Participation is encouraged for grid middleware and healthgrid application developers, biomedical and health informatics users, ethicists and security experts, and policy makers to participate in a set of multidisciplinary sessions with a common focus on bringing healthgrids closer to real application in the health domain.

HealthGrid conferences have been organized on an annual basis. The first conference, held in 2003 in Lyon (http://lyon2003.healthgrid.org), reflected the need to involve all actors – physicians, scientists and technologists – who might play a role in the application of grid technology to health, whether healthcare or bio-medical research.

The second conference, held in Clermont-Ferrand in January 2004 (http://clermont2004.healthgrid.org) reported on the earliest efforts in research, mainly work in progress, from a large number of projects. The third conference in Oxford (http://oxford2005.healthgrid.org) had a major focus on first results and exploration of deployment strategies in healthcare. The fourth conference in Valencia (http://valencia2006.healthgrid.org) aimed at consolidating the collaboration among biologists, healthcare professionals and grid technology experts. The fifth conference in Geneva (http://geneva2007.healthgrid.org) focused on the five domains defined by the European Commission as application areas for 'vertical integration' through grids in the biomedical field: molecules, cells, organs, individuals, and populations. For each of these five domains, an invited speaker gave a state of the art address followed by concrete projects. This was a loud signal to the community that the usefulness of grids to potential application domains could be demonstrated at least at the prototype level. This theme was also evident at the sixth conference in Chicago (http://chicago2008.healthgrid.org), which proclaimed its focus as 'e-Science Meets Biomedical Informatics'. The sixth conference was also a landmark in the history of the organization HealthGrid – and its newly established affiliate HealthGrid.US – as the first conference to be organized outside Europe. As we put it at the time, this was a celebration of similarities and differences, a moment to validate models and principles beyond one's familiar shores.

[1] Corresponding Author: Tristan Glatard (glatard@creatis.insa-lyon.fr)

The seventh conference returned to Europe, taking place for the first time in Germany in Berlin (http://berlin2009.healthgrid.org). While most themes touched on in earlier conferences continued to be present, certain other themes came to the fore perhaps more clearly than ever before: accessibility, the fraught challenge of usability, and the question of a business case for healthgrids. Ethical, legal, social and economic issues have now been encapsulated in the acronym ELSE, when it is no longer a joke that they are "someone else's problem". Perhaps more clearly than before, the Berlin conference embraced and debated these issues directly. An emphasis on cloud computing was also felt for the first time. 'Cloud' – in some form or other – began to be more obviously considered at the next conference in Orsay, near Paris, in 2010 (http://paris2010.healthgrid.org/). Certainly, the themes – *Socio-Economic Aspects and Accessibility*; *Future of Grids, Core Technologies and Data Integration*; and *Applications* – reflected a settled research agenda, but there was a great deal of innovation under discussion both on the technological and the implementation and deployment fronts.

Last year's conference in Bristol, UK, (http://bristol2011.healthgrid.org/) was the first to be co-located with another conference, on this occasion CBMS 2011, or as it is known formally, the Twenty-fourth International Symposium on Computer-Based Medical Systems. While there was an undeniable interest in clouds and their evolution – and differences – from grids, there was also, perhaps thanks to the CBMS link, a significant emphasis on medicine and healthcare.

While the desire to promote healthgrid applications to real healthcare settings remains central to the community's ambitions, it is also true that the majority of adopters work in academic environments where research is their principal preoccupation. In 2010 we wrote

> *Judging from the full range of papers submitted this year, cloud computing appears set to make an impact, just when healthcare informatics appears readier to adopt that paradigm, perhaps in preference to grids. So it may turn out that grids will remain the infrastructure of choice for research and clouds for (the business of) healthcare.*

Reality today has in fact proved rather more interesting, as cloud increasingly provides the focal model of distributed 'enterprise' computing but the potential of clouds to support grids, both to complement and to supplement them.

Conference in Amsterdam, 2012

The call for contributions resulted in 27 submissions from nine countries, among which 17 were full or short papers, three were tutorials, and seven proposed system demonstrations. All papers were read by three independent reviewers, and at least by two reviewers for tutorials and demonstrations. In total, nine full papers, two tutorials, and six demonstrations were accepted for presentation at the conference[2].

[2]Two demonstration abstracts do not appear in these proceedings because the authors were not able to attend the conference.

Part One – HealthGrid Applications and Technologies

A clear trend towards the discussion of different issues related to data management and sharing emerges from the contributions. Price *et al.* present a scalability analysis of federated queries expressed with the *Distributed caGrid Query Language* on a cohort of 2,972,969 patients. Results show linear speed-up with respect to the number of storage nodes, and thus provide evidence for the scalability of the process. Two other papers focus on the sharing of medical imaging data. Chervenak *et al.* target the sharing of brain scans and associated health information. They present and deploy a complete architecture encompassing uniformed image acquisition protocols, secured sharing across research institutions, distributed image retrieval, and the handling of patient consent. Schuler *et al.* describe a system to share collections of virtual slides and associated metadata among research pathologists, clinical pathologists, and scientists. Their system is deployed and actively tested in production conditions at two medical centres. At a more general level, Ainsworth *et al.* propose to use so-called *Research Objects* for the sharing of health data, knowledge, and expertise. The resulting eLab system was deployed for applications in primary care, long-term conditions management, bariatric surgery and public health. Finally, Mouw *et al.* present a study of the legal constraints of genetic data processing in European grids. Based on EU regulations on genetic material, current characteristics of grid architectures, and related cases, they draw conclusions on the way to use and design grid systems for the analysis of genetic data.

Supporting computational requirements of applications is still an active area. Moretti *et al.* present the *gcodeml* tool to analyse large phylogenetic datasets both on grids and computational clusters. Their system is ready for production processing of popular databases. Carrión *et al.* show a cloud-supported service for the BLAST application. The client is multi-platform and uses the same interface as the sequential tool to reduce the learning curve and integration effort.

The last two full papers belong to the life-science grid community session (http://lsgc.org). Michel *et al.* describe the activities of the technical team supporting the *biomed* virtual organization in using the European Grid Infrastructure. Necessary tooling, daily tasks, and procedures are described, and the paper concludes with suggestions about ways to decrease the human cost of this activity. Madougou *et al.*, from the *vlemed* virtual organization, describe a provenance information system integrated in the MOTEUR workflow manager. It captures data provenance from the execution logs of distributed biomedical applications and it offers a user-friendly web interface for navigation through the provenance data.

We conclude with abstracts describing the accepted tutorials and demonstrations. Visser describes a hands-on session with the high-performance computing cloud for interactive R-cluster and several other applications. Wu *et al.* propose a comprehensive hands-on session where the participants can develop their own application-specific grid portlet based on solutions developed within the German D-Grid/NGI-DE. Three demonstrations are related to neuroimaging applications. Wu *et al.* focus on interactive vizualisation using a local (browser) or remote (cluster-based) solution. Galeazzi *et al.* demonstrate the DECIDE platform for diagnosis and research on Alzheimer's disease. Korkhov *et al.* present the SHIWA platform for exchanging and combining interoperable neuroimaging workflows between three infrastructures. Finally, Torterolo and Ruffino demonstrate the Desktop Cloud Visualization for remote access of 2D and

3D interactive applications by geographically dispersed doctors for collaborative image sharing.

Part Three: Science Gateways for Biomedical Research

Science gateways for biomedical research combine the research on science gateways with use cases in HealthGrid. Shahand *et al.* have worked on the integrated support of imaging techniques in the field of neurosciences. They developed a medical image analysis pipeline facilitating the handling of vast amounts of data acquired with modern imaging techniques. The study presented this paper discusses the multiple phases of data processing leading to a virtual laboratory for medical image processing. Rouilly *et al.* process image-based RNAi (iBRAIN2) analysis. The research presented here focuses on the correlation of cellular phenotypes with specific environmental or internal challenges. The analysis, curation, and comparison of datasets from genome-wide screens are illustrated. They introduce to the requirements for the iBRAIN2 system that was developed to handle the wealth of acquired data.

The program in the tenth HealthGrid conference in 2012 shows a clear focus on data, rather than on processing. This shift follows trends also observed in many other scientific areas, which can be interpreted as a sign of maturity of the community, which is now starting to explore the value of the available data. We hope that the presented material will be insightful and suggestive of new perspectives for your research of infrastructures and technology for healthcare.