Workshop GARR "Calcolo e Storage distribuito", Rome, 29-30 November 2012

Progetto DECIDE: un esempio d'infrastruttura al servizio della comunità biomedica

V. Ardizzone[1)], R. Barbera[2),3)], A. Calanducci[4)], M. Fargetta[4)], F. Galeazzi[1)], I. Porro[5)], G. La Rocca[2)], S. Monforte[2)], R. Ricceri[2)], R. Rotondo[1)], D. Scardaci[2)], A. Schenone[5)] on behalf of the DECIDE Consortium

[1)] Consortium GARR, Via dei Tizii, 6, 00185 Rome – Italy, [2)]INFN, Division of Catania, Via S. Sofia, 64, 95123 Catania – Italy, [3)] Department of Physics and Astronomy of the University of Catania, Viale A. Doria, 6, 95125 Catania – Italy, [4)] Consorzio COMETA – Via S. Sofia, 64 – 95123 Catania – Italy, [5)] Department of Communication, System and Computer Science, University of Genoa, Viale Causa 13 - 16145 Genoa - Italy

Abstract

Alzheimer's disease is the most common cause of dementia (40-70% of the cases, in population over 65 years). Due to the increase in life expectancy and the disease's slow progression, the prevalence of dementia is going to increase dramatically over the next few decades. Dementia has an enormous societal impact on families, governments and their health and social sectors, the relevant cost being estimated to be globally comparable to the world's 21st largest economy. Alzheimer's disease is not only a European priority but rather a planetary one. In this framework, the way researchers look at Alzheimer's and dementia is changing considerably, especially after the development of new diagnostic guidelines in 2010-2011 by the National Institute on Ageing and the Alzheimer's Association. Under the new guidelines dementia is seen as a late stage in the evolution of the disease, while the guidelines aim to identify AD at an early stage of development by using a combination of results from structural (MRI), functional (FDG-PET) and molecular (PiB-PET) imaging as well as biochemical tests (analysis of CSF). In addition, the scientific community increasingly supports the idea that analysis of electroencephalography tests (EEG), in terms of power density and spectral coherence, although currently providing only moderate sensitivity and specificity may be used for the preliminary screening of large population samples due to its simplicity, low cost and non-invasiveness.

The DECIDE project is focused on supporting neurologists and physicians involved in the assessment of neurodegenerative diseases in the diagnosis and prognosis and aims at enhancing user confidence by improving the reliability of the required analysis and by integrating different clinical approaches. It provides a service, relying on a Grid e-Infrastructure, not aiming at the statistical combination of different reports, and its targeted community is the clinical one more than the research one. Mainly goal is to provide doctors at peripheral hospitals with service tools for determining clinical markers for the early diagnosis of neurological and psychiatric disorders (neurodegenerative diseases like Alzheimer and schizophrenia) together with their prognostic relevance. Thus the objectives and most notably the methodology differ significantly, requiring the development of a new and innovative infrastructure.

The project's aim is thus more on the clinical community, and this brings specific constraints in terms of ease of use, standardization, security, data confidentiality, data distribution. The service offered to such community would need to be compatible with the clinical routine, and thus be simple to use, robust, reasonably fast and should not require too much interaction with the system: of course, it should have also been validated with real cases beforehand, to ensure accurateness. Moreover, the ethical and legal issues related to data handling, processing and distribution need to be specifically addressed. The project aims to become a reference in the landscape of similar initiatives at European level, which is often characterized by duplication and fragmentation of effort. For this reason, the service would have to be flexible (to accommodate different needs) and extensible (to allow for the addition of new tools). The DECIDE platform consists of three different layers: research networks, Grid resources and domain-specific applications.

- Network connectivity, provided by the GÉANT backbone and the National Research and Education Networks of the countries participating in the project, brings together different types of structures (clinical and research centres and academic research institutions) interconnected among them by high speed and huge bandwidth links.
- The Grid infrastructure is used as a platform to enable collaboration among all partners, as a technological "glue" to harmonize and unify developments, and as an elastic pool of computing and storage resources where large volumes of data can be hosted and related analyses can be performed.
- The approach adopted by DECIDE regarding its applications ensures the requirements of the neurological community are taken into account from the very beginning in the design of application services to reach full usability in a real clinical environment. The use of different medical acquisition

data (Magnetic Resonance Imaging - MRI, Positron Emission Tomography - PET, and Electroencephalography - EEG) allows the combination of complementary diagnostic approaches on neurodegenerative disease diagnosis, enabling synergies between different clinical domains and possibly supporting correlation studies among different neurological approaches.

The inspiring idea behind the project has not been that of doing something "for" a community, but rather "together with" a community. Users of the service have been involved very early in, and all along, the development phase of the service, to ensure their requirements would be taken into account, and the resulting service would be accepted and usable. A key decision which was taken early on in the project, was that of adopting standards wherever possible. Adoption of standards helps to reduce development and maintenance costs, and is an important driver to simplifying service adoption and enlarging the user community. The DECIDE service is based on standards at all levels. From ICT perspective, this is true at network and Grid middleware levels (EMI/gLite, adopted on official production sites of EGI, the European Grid Infrastructure), up to Science Gateway design (JSR 168/286 for portlets, SAML for authentication, LDAP for users database, PKCS#11 for cryptography and SAGA for interface to Grid middleware). From the clinical perspective, the project has documented, and made publicly available, procedures for patient preparation, exam preparation, data acquisition and data quality control, with the twofold goal of improving the quality and the informative content of the acquired data and ensuring such data is consistent with the project's reference datasets. From the technological point of view, it was decided to develop and implement the service based on a Grid middleware, taking advantage of two features:

1. Availability of tools to integrate geographically distributed resources, in this case, by "resources" we mean chiefly databases of images, like those required to train algorithms for computing the volume of brain regions from MRI images, or those required to make statistical comparisons as in the case of FDG-PET images from "normal" cases.

2. Capability of establishing extremely fine-grained authorization policies, down to the level of the single user: this is a requisite in at least two respects: it allows owners of data to keep control on the subjects data is shared with, and it allows to control precisely who may have access to a given application.

On the other hand, using Grid technologies has some disadvantages for users, which would prevent adoption of the service in a clinical environment: specifically, coping with personal certificates and security procedures, execution scripts, job description languages and command-line interfaces tend to represent an obstacle to the adoption of the technology, especially by non-IT experts. All of these problems were solved by developing a Science Gateway (SG), namely a portal integrating a set of tools and applications, customized to meet the needs of the community: the problem of identifying users without relying on personal Grid certificates was solved by integrating into the SG support for robot certificates (which are Grid certificates used to identify the responsible for the unattended use of a service) and mediating access to them via identity federations.

The Science Gateway is an extremely powerful tool to make the Grid usable by user communities. It is based on the Liferay framework and portlet container, and fully supports the "portlet 2.0" standard JSR-286. Within such framework, applications are built and customized, according to the users' needs, by expert software developers by combining, whenever possible, existing portlets or writing new ones. Interaction with Grid services is mediated by the Grid Engine, a software layer compliant with the SAGA standard, specifically with its JAVA implementation (JSAGA), designed to interoperate with a number of middlewares. The Grid Engine effectively isolates the applications from the underlying stacks: indeed, the Science Gateway can successfully submit jobs to several infrastructures based on different middleware like gLite, Unicore, Globus (in use at EGI), OurGrid (Brasil), CNGrid (China) and Garuda (India). This isolation would make the possible transition to other computing paradigms (e.g., cloud) rather straightforward.

The main lesson learned is that prospected users' involvement, already since the design phase, is crucial to identify the specific needs and requirements of the targeted communities. At a later stage, cultivating a community of motivated early adopters and key opinion leaders ensures the service provided will be acceptable to the wider community. Moreover, sustainability issues have to be seriously considered, since the temporal horizon of a hospital or a research centre is typically of several years: in this respect, the choice of adhering to standards and of developing the service using a layered approach ensure the resulting product may be easily adapted to future new and evolving technologies.

We believe the DECIDE approach will play a significant role in bringing scientifically advanced and high quality clinical procedures to European citizens.