*eChallenges e-2012 Conference Proceedings*
*Paul Cunningham and Miriam Cunningham (Eds)*
*IIMC International Information Management Corporation, 2012*
*ISBN: 978-1-905824-35-9*

# The DECIDE Project: Designing and Implementing a Prototype Service for Supporting Early Diagnosis of Alzheimer's Disease

Fulvio GALEAZZI[1], Valeria ARDIZZONE[1], Mario REALE[1], Maddalena VARIO[1], Federica TANLONGO[1], on behalf of the DECIDE Consortium
[1]*Consortium GARR, via dei Tizii 6, Rome, 00185 Italy*
*Tel: +39 06 49622000, Fax: +39 06 49622044, Email:info@garr.it*

**Abstract:** This paper will present the design and implementation challenges of the innovative DECIDE [1] service, to support research and early diagnosis of Alzheimer's and other neurodegenerative diseases. DECIDE service, which is based on a Grid eInfrastructure, offers a set of tools providing quantitative measurements, to help researchers and clinicians make more informed diagnosis. As the service specifically targets the clinical community, it differs significantly from other initiatives since it needs to comply with the requirements imposed by the clinical routine in terms of accuracy, robustness, ease of use, data handling policies, adherence to clinical praxis. Moreover, sustainability aspects will also be discussed, since DECIDE aims to propose such service as a reference at European level, possibly extending it to other pathologies. We will then summarize the main results obtained to date, and the possible future developments.

## 1. Introduction

### *1.1 Framework*

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.

Indeed, a recent report by World Health Organization [2] estimated the number of people living with dementia worldwide at about 35 million, with this number due to nearly triple by 2050. Noteworthy, the fraction of such people living in low or middle income countries will rise, over the same time span, from 50% to 70%. Moreover, 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 Aging 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.

### *1.2 Motivation*

Among the recent and on-going initiatives related to Alzheimer's disease, in the category of projects not directly related to the testing or development of new drugs, the two most outstanding examples are neuGRID [3] (and its siblings outGRID and Neugrid4you) and PredictAD [4].

The neuGRID and related projects aim at providing researchers with a user-friendly Grid-based e-Infrastructure, and to develop a set of services in order to enable the European neuroscience community to carry out research for the study of degenerative brain diseases. Although generic in this approach, the projects are currently focusing on the management and analysis of AD data collected by ADNI and European-ADNI. The siblings projects outGRID and Neugrid4you aim at establishing interoperability with similar initiatives in US (LONI at UCLA) and Canada (CBRAIN at McGill), and at expanding the services offered and outreaching to new user communities, respectively.

On the other hand, PredictAD project's main focus is on combining data from various sources and extract an indicator to diagnose Alzheimer's disease as early as possible.

The DECIDE project is based on a new approach and differs significantly from both. It provides a service, relying on a Grid eInfrastructure, not aiming at the statistical combination of different reports, and its targeted community is the clinical one more than the research one. Thus the objectives and most notably the methodology differ significantly, requiring the development of a new and innovative infrastructure.

## 2. Objectives

The main long-term objective of the project is providing European citizens with high quality early diagnostic and prognostic procedures, irrespective of their financial or social status and location.

The project's focus 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). In this respect, the project is also focused on ensuring the sustainability of the service, the training of users and the extension to new communities (and possibly, new pathologies).

## 3. Challenges and Methodology

The inspiring idea behind the project has not been that of doing something "for" a community, but rather "together with" a community. The 13 partners making up the consortium represent all involved stakeholders: from network operators, Grid developers, GUI developers, representatives from the European network of excellence of research and healthcare centres specialized on AD, as well as patient's advocate association. Furthermore, 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 web front-end design (JSR 168/286 for portlets, SAML for authentication, LDAP for users database, PKCS#11 for cryptography and SAGA [5] for interface to Grid middleware). From the clinical perspective, the project has documented, and made publicly available [6], 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 [7] (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.

This setup allows for secure identification of users and automatically solves user management issues related to the digital identity lifecycle (for example, a retiring user is automatically banned from using the DECIDE service) but also creates a new issue, by introducing a N-1 mapping between users and robot certificates, which would make the service not compliant with the strict security procedures in use at any distributed computing infrastructure: the technical solution will be described in the next paragraph. It is important to note that over the last couple of years the effectiveness of the SG approach has been widely recognized, as is witnessed by the fact that more than 10 European Certification Authorities updated their procedures to issue robot certificates, and by the existence of numerous (~30 within EGI alone) SGs serving various communities.

Users of the service were classified into three groups, according to the functionality made available by the system:

- "Neurologists": these professionals take care of patients during the whole diagnostic process, from diagnosis to therapy. They need just to request exams to be performed by other service users and retrieve the relevant reports, which they will combine and use to make a diagnosis.

- "Physicians": these professionals (Radiologists, Neurophysiologists, Nuclear Medicine physicians) provide diagnostic information to Neurologists, for the specific test of competence.
- "Scientists": these users may have different scientific profiles (Physicists, Mathematicians, Statisticians,…). They deal with the diagnostic algorithms and collaborate with physicians by providing knowledge and comprehension of the underlying methodology.

These groups of users, or "roles", also correspond to an increasing degree of interaction with the service and have access to a role-customized version of the portal: they also require specific, tailored training.

## 4. Technology Description

As mentioned in the previous paragraph, the current implementation of the service is based on Grid middleware: as a first step, the Virtual Organization vo.eu-decide.eu was registered in the EGI Operations Portal, and the project's sites were configured to support such VO and offer a minimal set of Grid services (topBDII, WMS, LFC, CE, SE, UI).

To ensure the Service Level implied by the clinical relevance of the service, for "Physician" use it was decided to rely exclusively on certified EGI production sites, as these committed to respect minimum levels of availability and reliability. Although the current minimum thresholds for a site's availability and reliability set within EGI (70% and 75%, respectively) may look inadequate to guarantee production-level standards, the overall figures for the whole infrastructure (supporting the DECIDE VO) are much better, due to the statistically uncorrelated nature of sites' faults and the automated resubmission mechanism provided by the WMS. On the other hand, "Scientists" may run on any site supporting the VO. The choice of sites for job execution is enabled by requiring a specific software tag, to be published by the VO Software Manager.

The Science Gateway is an extremely powerful tool to make the Grid usable by user communities which may be sparse (not organized in a Virtual Research Community), non-IT expert, and whose main interest is in routinely using, rather than developing, a stable workflow. It is based on the Liferay framework and portlet container [8], 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. This modularity simplifies reusability of software, and eventually results in a shorter time-to-operation for new applications. As an example, an application dedicated to the computation of brain volumetry from MRI images, not originally foreseen by the project's Description of Work, was recently integrated in the SG, per the request of the project's Scientific Board: this required 3.5 man/months over a period of nearly 2 months. 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.

In the spirit of providing a service suitable for exploitation in the clinical practice, applications are designed keeping simplicity and ease-of-use in mind. Interaction with the user is kept to a minimum, to ensure the clinician finds an environment which closely resembles the one she is currently used to: for example, Physicians are generally required just to fill in the sex and age of the patient, upload the relevant data (PET, MRI, EEG) and then press the "Submit" button on the SG interface. Noteworthy, such simplicity has a

counterpart in the effort invested beforehand by developers in tuning the parameters of the involved algorithms, creating a workflow suitable to be executed unattended and capable of properly dealing with most common minor anomalies, and more generally in making sure the application will run in a reasonable amount of time, yet provide stable and accurate results. In no way would a clinician use a service frequently requiring manual resubmission of a failed job, nor would she (be allowed to) repeat an MRI or PET scan to get an image exactly matching the quality criteria expected by the applications.

To shield users from the hassle of requesting and managing personal certificates, the Science Gateway interacts, via a lightweight crypto-library, with the so-called eToken server: such server holds the robot certificates (one for each application/role, stored on a USB smartcard physically plugged in the server) and manages the creation of proxies on behalf of the user. Since, with this design, each user is funnelled through the same kind of proxy, the association of any Grid activity to the digital identity of the user is achieved owing to the Users Tracking DB (which is part of the Grid Engine), tracing and logging users' interaction with the Grid: this makes the Science Gateway compliant with EGI security procedures. As for the security aspects, the eToken server belongs to a private network and is only accessible from the SG: moreover, unlike other commonly used Science Gateways, the digital certificates never travel on the network, only the short-lived (12 hours, at most) proxies do. These features make security breaches very unlikely: however, in the worst possible case it will suffice to physically remove the USB smartcard to immediately (modulo the proxy lifetime) stop any unlawful Grid activity.

The SG decouples user Authentication from Authorization. As for Authentication, the SG supports several Identity Federations, like eduGAIN and IDEM [9]. A catch-all Identity Federation, maintained by the project, allows registration of users not belonging to any other federation: this is essential for capturing the targeted user communities, which are normally sparse, may not have a reference Identity Federation and may have little interest in setting up one to the only purpose of accessing the DECIDE service. The SG may be easily extended to other Identity Federations supporting the SAML2 standard in its Shibboleth and SimpleSAMLphp implementations [10]. Support for Identity Federations is a strong driver for the widespread adoption of the service, as new Identity Providers joining a Federation immediately translate into new potential users of the DECIDE portal.

Having ascertained the identity of a user, does not automatically imply such user is authorized to access DECIDE services. Authorization is a two-step procedure, supervised by the DECIDE Scientific Board. First, users must register to the site, which triggers the creation of an entry in an LDAP database, and then they need to attend application- and role-specific training course(s) and get the relevant qualification(s): only at this point their entry in the LDAP database is granted the role(s) matching the obtained qualification(s).
From the point of view of the SG, the benefits of this Authentication and Authorization mechanism are:

- Offloading of the identity management: no need to maintain (protect, update, verify validity of) user credentials on the SG
- Enforcement of procedures to ensure all users of the service have been previously properly trained
- Full control over who is authorized to use the service: for example, in case of a security breach a user may be banned by simply unchecking her roles in the LDAP database.

Since the infrastructure is managing patients' data, a lot of attention has been put on security issues and this, as was explained, was one of the reasons for choosing to develop basing on a Grid middleware. One kind of data is particularly critical, namely the FDG-PET images making up the reference dataset for normal cases. Unlike MRI images, in fact, no public databases exist storing such kind of images, and strong legal, ethical (and financial) constraints severely limit the possibility for hospitals and research centres of acquiring

them[1]. One of the applications offered by the DECIDE service performs a comparison of the patient's image with an "average brain" constructed by combining some number of normal subjects' images, to highlight regions of statistically significant glucose hypo metabolism[2]. The best results are obtained when using 50-100 normal images, but the normal dataset is being enlarged in order to allow filtering on sex or age or scanner manufacturer/model. Normal subjects' images are an immense wealth for a hospital: a large one may manage to get a mere order of ten such images per year.

To effectively share such precious data resources without giving up ownership, DECIDE adopted the SecureStorage middleware component [11], a Grid-aware solution to store encrypted data on Storage Elements: in this way, not even system administrators of the storage elements can access the confidential data in a clear format. Confidential data files are encrypted by a KeyStore (KS) server housed in the administrative domain of the data owner and then copied to one or more Storage Elements like any regular Grid file: application jobs will contact the KS to retrieve the decryption key and, if authorized to do so, will manage to access the data.

Information relevant to each data file is stored in an AMGA metadata catalogue: this, however, is never accessed directly but rather using gLibrary [12], a software package providing simplified access to and management of information for users and applications.

## 5. Results

At present date, the project has achieved all key mid-term objectives, specifically:

- All standards relevant to the different layers of the infrastructure (i.e., grid infrastructure and middleware, patient preparation, test execution, clinical practice) were identified, described and adopted
- The Grid-based infrastructure is up and running, all needed services are installed at EGI official, certified, production sites, including some (AMGA, gLibrary, SecureStorage) specifically setup for the DECIDE VO
- The reference databases have been designed and implemented, and populated with multi-modal data for patients of different classes: elderly normal subjects, mild cognitive impairment (MCI) converter and non converter, Alzheimer's disease
- The Science Gateway was deployed and customized, and counts more than 60 users to date, for all application and roles, removing overlaps. It is important to note, in this respect, that the DECIDE service addresses a very specialised audience and that, in order to exploit it, it is necessary for users to undergo a training and a final examination that certifies their understanding of the usage of the tool and of its outcome. This implies that many of the single users currently registered represent the whole department of neuroimaging at their hospital
- Five applications are available, for the analysis of PET images (brain functionality), EEG traces (neuro-physiology) and MRI images (brain volumetry)
- A training campaign is on-going, and dissemination activities led to presentation of the project to a number of stakeholders and key opinion leaders, generating expressions of interest and requests for involvement in the validation and usability test activities.

Overall, the infrastructure and service were found very useful by users, and their feedback (on the content of reports, on the way results should be presented and on the general guidance offered by the site) has been taken into account for improving what currently provided.

[1] "Normal" cases FDG-PET images are typically acquired from patients suffering from pathologies with no cerebral implication undergoing a total-body PET scan, by asking their explicit consent for also scanning the whole head region.

[2] All images (patient and normal) are co-registered to the same brain atlas before the comparison.

## 6. Developments

For the immediate future, the project is mainly focused on completing the foreseen activities and achieving its goals, especially in terms of users training and the setup of a strategy for sustainability: this latter issue is particularly critical and has implications at many levels.

As far as the infrastructure is concerned, we are leveraging our MoUs with infrastructure developers (EMI) and providers (EGI) to ensure commitment for future development and maintenance of services like gLibrary and the SecureStorage, and their deployment at an increasing number of official, certified sites.

Further extensions to the application portfolio are not foreseen through the end of the project due to the lack of time, within such timespan, to carry on an adequate training programme. In the longer term, however, this is one of our priorities, with the aims of providing a service which is more and more useful to our users and of attracting new ones, which are both key ingredients in the recipe for sustainability. Along the same line we will exploit the flexibility of the infrastructure, to extend it along three dimensions:

- Widening the application portfolio, including its extension to other pathologies, for example in the field of cardiology or oncology
- Adding support for combining information from different reports: currently, all exams relevant to a patient are stored in the same "diagnostic session" (sort of a virtual folder), but no tool is provided at present to correlate the various pieces of information
- Adding support for time-series, providing the possibility to compare the same kind of exams repeated after some time, for example after a period of rehabilitation, or to monitor series of clinical trials.

In part, this extension is already taking place, within the framework of new proposals which have been presented by partners of the project at national and European level, some of which have already been granted funding.

In addition to enlarging the user community and making the service more functionally attractive, a broad set of actions is being taken to contact trans-national initiatives like the Joint Programme initiative on Neurodegenerative Research (JPND), Health Ministries and other local bodies with competencies in the public health, professionals associations, and patients' advocate associations: ways to approach such entities range from the organization of bilateral meetings to facilitate personal contact, to presentation of the project at workshops and conferences, to dedicated communication via e-mail or newsletters. In this way we aim to find and exploit new channels for dissemination and outreach to the medical community, to spot possible strategic partnerships, and to identify sources of complementary funding after the end of the FP7 project lifetime.

The key messages to be emphasized are that the service is in line with the most recent scientific trends, that it has been validated on real cases (as per relevant high-level scientific publications) and that DECIDE approach, far from being a substitute for the clinician's judgment and experience, can dramatically change the way diagnoses are drawn. For example, the patient's reported subjective or objective deficit of cognitive functions may not be such to allow a clinician to draw a conclusion, even with the support of diagnostic imaging tests: in such cases the patient would need to repeat the test after some time, typically six months or one year. On the other hand, the DECIDE service may provide the additional quantitative information (invisible or not easily visible to the naked eye) which would help the clinician draw a definite diagnosis, (negative, or of Alzheimer's disease or of other forms or dementia) months before what is currently possible: this would benefit the patient (fewer invasive tests and, in the worst case, more time to plan ahead for the later stages of the disease) and the society at large (economic savings, possible development and test of new drugs to slow the progression of the disease).

The project is developing a business model and is currently performing a market analysis which will identify the main market sectors to target and the best strategy to tackle them. Before going to the market the project needs also addressing issues like, for example, protection of intellectual property rights, licensing schema for the applications and data accession rights after project end. As for the possible revenue streams, several possibilities are being investigated, including exploiting the training programme (possibly in connection with continuing medical education), certifying the service as a medical device, or offering it as a second-opinion service.

## 7. Conclusions

The challenges, the methodology used and the results obtained so far while developing a service of clinical relevance were described.

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.

User acceptability has an impact on the technologies adopted, and these must be explained and documented in detail so that users may build confidence: specifically, with respect to data ownership and data management (legal, ethical) issues. 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.

In parallel to the development of the service, an intense dissemination activity at all levels, supported by scientific evidence, is also needed, to propose success stories which strengthen the perceived usefulness of the service, conveying the message that the service is in line with recent clinical developments and complies with the clinical praxis.

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

## Acknowledgement

This work has been supported by the EC within the 7th Framework Programme under Grant Agreement n. 261593.

## References

[1] DECIDE Project: https://www.eu-decide.eu/
[2] World Health Organization report: http://www.who.int/mental_health/publications/dementia_report_2012/en/
[3] neuGRID Project: http://www.neugrid.eu/
[4] PredictAD Project: http://www.predictad.eu/
[5] SAGA OGF Standard Specification: http://www.gridforum.org/documents/GFD.90.pdf
[6] DECIDE Deliverables are available at: https://www.eu-decide.eu/index.php/documents/deliverables
[7] Science Gateways are discussed in: Wilkins-Diehr N., Gannon D., Klimeck G., Oster S., Pamidighantam S. (2008), TeraGrid Science Gateways and Their Impact on Science, IEEE Computer 41(11), 32-41.
[8] Liferay portal framework: http://www.liferay.com/
[9] Identity federations: eduGAIN, http://www.edugain.org/ ; IDEM, http://www.idem.garr.it/
[10] For more information: Shibboleth, http://shibboleth.net/ ; SimpleSAMLphp, http://simplesamlphp.org/
[11] Scardaci D., Scuderi G. (2007), A Secure Storage Service for the gLite Middleware, Proceedings of the Third International Symposium on Information Assurance and Security, p. 261-266.
[12] Calanducci A et al. (2007), A Digital Library Management System for the Grid, Fourth International Workshop on Emerging Technologies for Next-generation GRID (ETNGRID 2007) at 16th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE-2007), GET/INT Paris, France, June 18-20, 2007.

 