Final Results from and Exploitation Plans for MammoGrid

Florida ESTRELLA, Chiara DEL FRATE, Jose GALVEZ, Tamas HAUER, David MANSET, Richard McCCLATCHHEY, Mohammed ODEH, Dmitry ROGULIN, Tony SOLOMONIDES and Ruth WARREN

CERN, 1211 Geneva 23, Switzerland

CCCS Research Centre, Univ. of West of England, Frenchay, Bristol BS16 1QY, UK

Istituto di Radiologia, Università di Udine, Italy

Maat GKnowledge, Toledo, Spain

Breast Care Unit, Addensbrooke Hospital, Cambridge, UK

Abstract. The MammoGrid project has delivered the first deployed instance of a HealthGrid for clinical mammography that spans national boundaries. During the last year the final MammoGrid prototype has undergone a series of rigorous tests undertaken by radiologists in the UK and Italy and this paper draws conclusions from those tests for the benefit of the HealthGrid community. In addition lessons learned during the lifetime of the project are detailed and recommendations drawn for future health applications using Grids. Following completion of the project plans have been put in place for the commercialization of (some of) the MammoGrid system and this is also reported in this article. Particular emphasis is placed on the issues surrounding the transition from collaborative research project to marketable product. This paper concludes by highlighting some of the potential areas of future development and research.

Keywords. Medical imaging, Grid application, deployment, exploitation and commercialization.

1 Introduction

The EU funded project Mammogrid set out to explore the following conjecture: Grid technology and standards have evolved to the point where a prototype federated database of mammograms might be constructed, based on centres in three European countries (UK, Italy, Switzerland). The project was conducted between September 2002 and August 2005 and comprised partners from Oxford, Bristol and Cambridge in the UK, CERN in Switzerland and Institutes in Pisa, Sassari and Udine in Italy. The project has developed a set of prototypes, there have been regular papers published by MammoGrid and the reader is referred to these for technical aspects of the project [1], [2], [3]. In the final year of the project a final prototype has been constructed and has been deployed to the University hospitals in Cambridge and Udine for proof-of-concept evaluation that would demonstrate the use of a Grid-based medical platform in
clinical tests. More specifically, Mammogrid set out to explore the following clinical issues through the delivery of the first Grid-based mammography platform:

- **Image standardisation**: The appearance of a mammogram is greatly affected by differences in image acquisition processes (machine type, filter, exposure time etc.). Such differences can impact significantly radiologist judgements (presence of microcalcifications, estimation of proportion of dense breast tissue). Ideally, the given image would be “standardised” by removing such anatomically irrelevant variations prior to adding it to the database. Mammogrid explored the possibility of standardising images using the Standard Mammogram Form (SMF) representation developed by Highnam and Brady [4].

- **Breast density as a risk factor**: It has been suggested, primarily by Boyd [5] and others that the amount/percentage of dense tissue in the breast is a major (perhaps, the major) risk factor for breast cancer (after taking due account of lifetime experiences). The SMF representation provides a number of measures of the amount/proportion of dense tissue, so the MammoGrid clinical partners at Cambridge and Udine hospitals sought to compare measurements of breast density as provided by SMF with the standard methods of visual assessment [6] and the automated 2D interactive computer programme available from Jaffe and Boyd [7].

- **Computer-aided detection of microcalcifications & masses**: Prior to Mammogrid, the Project partners from Sassari and Pisa Universities had developed a system named CALMA (Computer Aided Library for Mammography) for the detection of lesions and microcalcifications with reportedly good sensitivity and specificity [8]. The aim was to use the database of mammograms generated during the Mammogrid project at Cambridge and Udine to (a) re-assess the sensitivity and specificity of CALMA, and (b) to examine whether its performance would be improved on the standardised images generated by SMF.

Mammogrid is one of a number of European HealthGrid projects e.g. [9], [10], [11]. We note also that the UK “e-Science” programme has funded the e-Diamond project [12], which also aims at developing a federated database of mammograms, but whereas Mammogrid is based on open-source software, eDiamond is based on (IBM) proprietary technology and concentrates on two complementary applications, namely teaching and FindOneLikeIt. Also, in the United States, the National Digital Mammography Archive (NDMA) [13] has adopted a radically different approach, using a large centralised archive.

In this paper we report on the outcome of the clinical tests in MammoGrid and identify how far the project progressed towards real clinical use of a healthgrid. The paper identifies both achievements and obstacles in the use of a deployed Grid-based healthcare application and it highlights the lessons learned from MammoGrid. In addition the exploitation plans for MammoGrid are outlined since post-project an initiative has begun for commercialisation of the project software.

### 2 The MammoGrid Clinical Evaluation

The Mammogrid project has several technical aspects, which are briefly mentioned in this section in order to convey a flavour of the scope and complexity of the challenges that were met. These aspects include: image standardisation using SMF; the development of a workstation on which images can be acquired, annotated,
and uploaded to the Grid; and the distribution of data, images and clinician queries across Grid-based databases, while respecting ethical, legal, confidentiality and security constraints applicable differently in the partner countries of origin. It was not the intention of this project to produce new Grid software. Rather, the aim has been to use, wherever possible, freely available software to enable radiologists to query images across a widely distributed “federated” database of mammographic images and to perform epidemiological and Computer-aided detection on the sets of returned images. For example, in the Mammogrid project, radiologists may annotate (i.e. mark out) different regions of a mammogram, which are then subjected to different computer-aided detection algorithms (including CALMA) and compared with stored mammograms in the database. Since any one of these stages may be executed independently or take some time to be completed, the process must be controlled in a way that recognises the current state of the computation and ensures that results are meaningfully assembled from the various partial outcomes. To provide for these possibilities, Mammogrid has built on the results of previous European Grid projects such as EGEE [14]. The details of EGEE and how it was used in the design of Mammogrid are beyond the scope of this paper.

The MammoGrid project has recently delivered its final proof-of-concept prototype enabling clinicians to store digitized mammograms along with appropriately anonymized patient metadata; the prototype provides controlled access to mammograms both locally and remotely stored. A typical database comprising several thousand mammograms has been created for user tests of clinicians’ queries. The prototype comprises

- a high-quality clinician visualization workstation (used for data acquisition and inspection);
- an imaging standard-compliant interface to a set of medical services (annotation, security, image analysis, data storage and querying services) residing on a so-called ‘Grid-box’; and
- secure access to a network of other Grid-boxes connected through Grids middleware.

![Figure 1: The MammoGrid Virtual Organisation](image)

To facilitate evaluation of the final prototype at the clinical sites, a MammoGrid Virtual Organization (MGVO) was established and deployed (as shown in figure 1). The MGVO is composed of three mammography centres – Addenbrookes Hospital, Udine Hospital, and Oxford University. These centres are autonomous and independent of each other with respect to their local data management and ownership. The Addenbrookes and Udine hospitals have locally managed databases of mammograms,
with several thousand cases between them. As part of the MGVO, registered clinicians have access to (suitably anonymized) mammograms, results, diagnosis and imaging software from other centres. Access is coordinated by the MGVO central node at CERN.

The adopted Grid implementation is the ALICE Environment (AliEn) [15] component of the EGEE-gLite middleware [16], the grid middleware of the EU-funded EGEE project. The service-oriented approach adopted in MammoGrid permits the interconnection of communicating entities, called services, which provide capabilities through exchange of messages. The services are “orchestrated” in terms of service interactions: how services are discovered, how they are invoked, what can be invoked, the sequence of service invocations, and who can execute them.

The MammoGrid Services (MGS) are a set of services for managing mammography images and associated patient data on the grid. Figure 2 illustrates the services that make up the MGVO. (For simplicity, Oxford University has not been included). The MGS are: (a) Add for uploading files (DICOM [17] images and structured reports) to the MGVO; (b) Retrieve for downloading files from the grid system; (c) Query for querying the federated database of mammograms; (d) AddAlgorithm for uploading executable code to the grid system; (e) ExecuteAlgorithm for executing grid-resident executable code on grid-resident files on the Grid system; and (f) Authenticate for logging into the MGVO. For further details consult [3].

Currently the MammoGrid Virtual Organisation encompasses data accessible to senior radiologists at the Addenbrookes and Udine hospitals, as well as at Oxford University and CERN. The radiologists have been able to view raw image data from each others hospitals and have been able to second read Grid-resident mammograms and to separately annotate the images for combined diagnosis. This has demonstrated the viability of distributed image analysis using the Grid and shown considerable

![Figure 2: The MammoGrid Services in the MGVO.](image)
promise for future health-based Grid applications. Despite the anticipated performance limitations that existing Grid software imposes on the system usage, the clinicians have been able to discover new ways to collaborate using the virtual organisation. These include the ability to perform queries over a virtual repository spanning data held in Addenbrookes and Udine.

Clinicians define their mammogram analysis in terms of queries they wish to be resolved across the collection of data repositories. Queries can be categorized into simple queries (mainly against associated data stored in the database as simple attributes) and complex queries which require derived data to be interrogated or an algorithm to be executed on a (sub-)set of distributed images. The important aspect is that image and data distribution are transparent for radiologists so queries are formulated and executed as if these images were locally resident. Queries are executed at the location where the relevant data resides, i.e. sub-queries are moved to the data, rather than large quantities of data being moved to the clinician, which can be prohibitively expensive given the quantities of data. Figure 3 illustrates how queries are handled in MammoGrid.

The Query Analyzer takes a formal query representation and decomposes it into (a) a formal query for local processing, and (b) a formal query for remote processing. It then forwards these decomposed queries to the Local Query Handler and the appropriate Remote Query Handler for the resolution of the request. The Local Query Handler generates query language statements (e.g. SQL) in the query language of the associated Local DB (e.g. MySQL). The result set is converted to XML and routed to the Result Handler. The Remote Query Handler is a portal for propagating queries and results between sites. This handler forwards the formal query for remote processing to the Query Analyzer of the remote site. The remote query result set is converted to XML and routed to the Result Handler. As of writing the MGVO holds:

<table>
<thead>
<tr>
<th>Site</th>
<th>Number of Patients</th>
<th>Number of Image Files</th>
<th>Number of SMF Files</th>
<th>Associated Database Size</th>
<th>File Storage Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge</td>
<td>813</td>
<td>2798</td>
<td>2738</td>
<td>29 Mb</td>
<td>70 Gb</td>
</tr>
<tr>
<td>Udine</td>
<td>489</td>
<td>4663</td>
<td>2372</td>
<td>37 Mb</td>
<td>85 Gb</td>
</tr>
</tbody>
</table>

Table 1: Virtual Repository size of the MammoGrid prototype.
The average processing time for the core services are: (1) Add 8Mb DICOM file takes approximately 7 seconds (2) Retrieve 8Mb DICOM file from a remote site takes approximately 14 seconds (3) SMF workflow of ExecuteAlgorithm and Add takes around 200 seconds. For querying:

<table>
<thead>
<tr>
<th>Query</th>
<th>Cambridge</th>
<th>Udine</th>
<th>Num images</th>
<th>Num patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Id: Cambridge patient</td>
<td>2.654 sec</td>
<td>2.563 sec</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>By Id: Udine patient</td>
<td>2.844</td>
<td>3.225</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>All female</td>
<td>103</td>
<td>91</td>
<td>12571</td>
<td>1510</td>
</tr>
<tr>
<td>Age [50,60] and ImageLaterality=L</td>
<td>19.489</td>
<td>22.673</td>
<td>1764</td>
<td>357</td>
</tr>
</tbody>
</table>

Table 2: Data query performance of the MammoGrid prototype.

Clinicians continue in the process of scanning and annotating cases contributing to several ongoing medical studies. These include (1) Cancers versus Control study: breast density study using SMF standard (2) Dose/Density study: exploring the relationship between mammographic density, age, breast size and radiation dose (3) CADe and validation of SMF in association with CADe. These studies aim to show how health professionals can work together without co-locating, and the MammoGrid approach provides a suitable environment for new forms of clinical collaboration.

In the final months of the project the clinicians have been testing the MammoGrid prototype across two applications. First, the project has facilitated the use of the Standard Mammogram Format (SMF) [4] software to measure breast density. The clinical project, designed jointly by Cambridge and Udine, explores the relationship between mammographic density, age, breast size, and radiation dose. In this project, breast density has been measured by SMF and compared with standard methods of visual assessment. Heights, weights, and mass indicators are used in an international comparison, but a richer dataset would be needed to study effects of lifestyle factors such as diet or HRT use between the two national populations. Second, the University of Udine will lead a project to validate the use of SMF in association with CADe from the CALMA project [18] & [19]. Cancers and benign lesions will be supplied from the clinical services of Udine and Cambridge to provide the benchmarking and the set of test cases. Cancer cases will include women whose unaffected breast will serve the density study to provide cases for the CADe analysis from the affected side mammogram. The MammoGrid project has shown that these new forms of clinical collaboration can be supported using the Grid (see [20]).

3 Lessons Learned

The nature of the project and its particular constraints of multi-disciplinarity, dispersed geographical development, large discrepancies in participants’ domain knowledge, whether of software engineering techniques or of breast cancer screening practice, and the novelty of the Grid environment provide experiences from which other Grid-based medical informatics projects can benefit. We summarise below some of the main lessons that can be learned in this context.

First, the project was particularly fortunate in its medical partners. In general, the medical environment is very risk-averse, conservative in nature and reluctant to adopt
new technologies without significant evidence of tangible benefit. It is therefore important to identify a suitable user community in which new technologies (such as Grid) can be evaluated. In the case of MammoGrid we have had real commitment from the radiology community in the project’s requirements definition and analysis, implementation, and evaluation and this was crucial to the success of the project. The data samples used were of a sensitive nature and required both ethical clearance from participating institutions and anonymization of the data and even then only for strictly research use in the project. Many ethical/legal obstacles remain to be tackled before clinicians can share sensitive patient data between institutes, never mind across national boundaries.

Second, it has become clear from our experiences that Grid middleware technology itself is still evolving, and this suggests that there is a clear need for standardization to enable production-quality systems to be developed. Despite the availability of toolsets such as the Globus 4.0 [21] the development of applications that harness the power of the Grid, as yet, requires specialist skills and is thus costly in terms of manpower. Only with the arrival of stable middleware and packaged Grid services will the development of medical applications become viable.

Third, the performance of existing middleware is also somewhat suspect; the MammoGrid project had to circumvent some of the delivered Grid services to ensure adequate system performance for its prototype evaluation. For example, the database of medical images was taken out of the Grid software to provide adequate response for MammoGrid query handling. The EGEE [14] project is addressing these technological deficiencies and improved performance of the middleware should consequently be delivered in the coming years.

Fourth, Grid technology for medical informatics is still in its infancy and needs some proven examples of its applicability; MammoGrid is the first such exemplar in practice. Equally, awareness of Grid technology and its potential (and current limitations) must still be raised in the target user communities such as Health, Biomedicine, and more generally life sciences.

Fifth, the project has indicated that it is possible to use modelling techniques (such as use-cases from UML) in a widely distributed, multi-disciplinary software engineering problem domain, provided a very pragmatic approach is used, where adopting a certain modelling technique is, to some extent, independent from the software development life cycle model being applied. The MammoGrid project has benefited significantly in its coordination, communication and commitment by utilizing the use-case model as the lingua franca during user requirement analysis and system design rather than following the disciplines of RUP to the letter.

Sixth, the evolutionary approach to system development work packages has mitigated the effects of the project constraints of a highly dynamic research-oriented environment in which novices and specialists in software engineering have worked together even though they may have been geographically separated.

Further areas that might promote the use of rigorous software engineering disciplines in the design of Grid-based software services are that of model-driven engineering [22] and the use of architecture descriptions [23] as the basis for the generation of Grid-wide services. These aspects are, however, outside the scope of the current project.
4 Future Exploitation Plans

By using Grid computing, the MammoGrid system allows hospitals, healthcare workers and researchers to share data and resources. It supports effective co-working, such as obtaining second opinion, provides the means for powerful comparative analysis of mammograms and opens the door to novel, broad-based statistical analysis of the incidence and forms of breast cancer. Through the MammoGrid project, partners have developed a strong collaboration between radiologists active in breast cancer research and academic computer scientists with expertise in applications of Grid computing. The success of the project has led to interest from outside companies and hospitals, with one Spanish SME, Maat GKnowledge, looking to deploy a commercial variant of the system in the region of Extremadura in Spain. Maat GKnowledge aims to have a MammoGrid system throughout the region that will enhance the ability of doctors to verify test results, to obtain a second opinion and to make use of the clinical experience acquired by the hospitals involved in the project. They then aim to scale it up in terms of what the system can do and the geographical area where it is used, expanding it to other areas of Spain and then Europe.

The MammoGrid system prototype has been at the leading edge of the healthgrid revolution and implements for the first time such a solution for mammogram acquisition and manipulation. The resulting application has reached a high level of complexity which now requires continued partnership between academics, clinicians and industry to provide the necessary technology transfer to enable real commercialisation. In this context, the MammoGrid Technology Transfer and Innovation eXchange (MaTTrIX) project aims to transfer project knowledge and expertise from the research to the commercial domain, to make its innovation available to the Maat GKnowledge company in Spain, to carry research findings into radiological practice and to reinforce existing partnerships between networks of clinicians and technologists. To achieve these objectives, the intention is to introduce a service which exploits the findings of the MammoGrid project in practice, while the researchers also learn from the application of their ideas in a real environment. This will create a two-way innovation flow between the academic and commercial worlds, building on existing synergies and collaborations, and improving the overall viability and commercialisation of the MammoGrid system software.

The MaTTrIX technology transfer approach relies on two main cornerstones:

- to promote the mobility of key experienced researchers, to absorb, expand and disseminate the knowledge for the MammoGrid system to evolve gradually into a commercial offering, making it available for healthcare across the European Research Area (ERA) as a viable and clinically assessed solution; and
- to create and to develop a strategic and durable partnership between new hospitals and partners of the MammoGrid project, enabling the constitution of a network of excellence in European research.

To this end, the host organisation Maat GKnowledge, the University of the West of England (UWE - Bristol), CIEMAT (Spain) and CERN together with clinical partners at the University hospitals in Cambridge (UCAM) and Udine (UNIUD) and at the Hospitals Infanta Cristina, Merida, San Benito in Extremadura are initiating a transfer of knowledge, research competences and technologies to enable the company to take over future development of the MammoGrid system. This will enable the training of existing Maat GKnowledge staff in MammoGrid technologies and the acquisition of the technology and know-how for commercialisation. This transfer
would constitute the mechanism for the creation and development of a durable
technology and knowledge transfer partnership.

While having paved the way for potential knowledge discovery in the
understanding of breast cancer, the MammoGrid project has also determined new
important research paths for better cancer prediction and diagnosis. The use of a
standard format for mammogram images (SMF) and its outcome in the
epidemiological investigations, has demonstrated the relevance of new Grid-based
clinical studies, which may lead to major advances in cancer prediction. In addition,
while most computer-aided detection (CADe) systems process raw and noisy data at
the price of accuracy and quality, the implemented solutions in MammoGrid have
indicated the valuable joint use of SMF and CADe tools to improve automated cancer
diagnosis. Considering the IT contribution in MammoGrid, the distributed technologies
used in the project combined with the recent clinical feedback, obtained from the
assessment of the proof-of-concept prototype, have highlighted the importance of
offering a collaborative platform for healthcare. Not only would such a solution
demonstrate the benefits of clinical second opinion, but it would also help in reducing
the information infrastructure costs by enabling heterogeneous and scalable resource
sharing, resulting in an improved system access including to less favoured regions and
countries in Europe.

Emphasising this last point the Hospitals of Infanta Cristina, Merida and San
Benito in Extremadura in Spain, will obtain access to the MammoGrid system and
expertise at a reduced cost, the infrastructure being already in place. While accessing
the latest technologies and software, these hospitals will share and enrich their clinical
experience by interacting with other trained clinicians from the University Hospitals in
Cambridge and Udine to obtain expertise in the use of SMF and CADe. This symbiosis
will result in improved and refined clinical knowledge being made available in a
Europe-wide production reference database, enabling new clinical studies in the spirit
of improving cancer prediction and detection.

5 Conclusions

The MammoGrid project has deployed its first prototype and has performed the
first phase of in-house tests, in which a representative set of mammograms have been
tested between sites in the UK, Switzerland and Italy. In the next phase of testing,
clinicians will be closely involved in performing tests and their feedback will be
reflectively utilised in improving the applicability and performance of the system. In its
first two years, the MammoGrid project has faced interesting challenges originating
from the interplay between medical and computer sciences, and has witnessed the
excitement of the user community whose expectations from a new paradigm are
understandably high. As the MammoGrid project moves into its final implementation
and testing phase, further challenges are anticipated. In conclusion, this paper has
outlined the MammoGrid application’s deployment strategy and experiences. Also, it
outlines the strategy being adopted for migration to the new middleware called gLite.

MammoGrid is one of several current projects, which aim to harness recent
technological advances to achieve the goal of complex data storage in support of
medical applications. The approaches vary widely, but at least two projects, e-
Diamond in the UK and MammoGrid in Europe, have adopted the Grid as their
platform of choice for the delivery of improved quality control, more accurate
diagnosis and statistically significant epidemiology. Since breast screening in the UK and in Italy has been based on film, mammograms have had to be digitised for use in both e-Diamond and MammoGrid. By contrast, the NDMA project in the United States has opted for centralised storage of direct digital mammograms.

The central feature of the MammoGrid project is a geographically distributed, Grid-based database of standardised images and associated patient data. The novelty of the MammoGrid approach lies in the application of Grid technology and in the provision of data and tools, which enable radiologists to compare new mammograms with existing ones on the Grid database, allowing them to make comparative diagnoses as well as judgements about quality. In the longer term, as the potential of the database is to be populated with provenance-controlled, reliable data from across Europe, with the prospect of statistically robust epidemiology that allows analysis of ‘lifestyle’ factors, including, e.g., diet exercise and exogenous hormone use. And, hence the Grid would also be suitable for storing genetic or pathological image information.

The project has attracted attention as a paradigm for Grid-based radiology and imaging applications. While it has not yet solved all problems, the project has established an approach and a prototype platform sharing medical data, especially images, across a Grid. In loose collaboration with a number of other European medical Grid projects, it is addressing the issues of informed consent and ethical approval, data protection, compliance with institutional, national and European regulations, and security [24], [25].

Acknowledgements

The authors thank the European Commission and their institutes for support and acknowledge the contribution of the following MammoGrid collaboration members: Mike Brady and Chris Tromans (Univ of Oxford), Predrag Buncic, Pablo Saiz (CERN/AliEn), Martin Cordell, Tom Reading and Ralph Highnam (Mirada) Piernicola Oliva (Univ of Sassari) and Evalina Fantacchi and Alexandra Rettico (Univ of Pisa). The assistance from the MammoGrid clinical community is warmly acknowledged especially that from Iqbal Warsi and Jane Ding of Addensbrooke Hospital, Cambridge, UK and Dr Massimo Bazzocchi of the Istituto di Radiologia at the Università di Udine, Italy. In addition the authors are indebted to the support and innovation of the former MammoGrid Project Coordinator, Roberto Amendolia.

References
