A relational approach to the capture of DICOM files for Grid-enabled medical imaging databases

David Power Eugenia Politou Mark Slaymaker Steve Harris Andrew Simpson
Oxford University Computing Laboratory (U.K)
{David.Power,Eugenia.Politou,Mark.Slaymaker,Steve.Harris,Andrew.Simpson}@comlab.ox.ac.uk

ABSTRACT
The Standard for Digital Imaging and Communications in Medicine (DICOM) specifies a non-proprietary digital imaging format, file structure and data interchange protocols for the transfer of biomedical images and non-image data related to such images—it is a specification of the components that are required in order to achieve inter-operability between biomedical imaging computer systems. In this paper we describe how a Grid-enabled medical imaging database—eDiaMoND—employs an object-relational approach to the storage of DICOM files. Although the work described has been carried out within the context of a particular mammography related project, the underlying principles are applicable to other medical imaging systems dealing either with other modalities or with other diseases.

Categories and Subject Descriptors
J.3 [**Life and Medical Sciences**]: Medical information systems; H.2.1 [**Database Management**]: Logical design; H.2.8 [**Database Management**]: Database Applications—*Image databases*

General Terms
Design, Reliability

Keywords
databases, medical imaging, DICOM, mammography

1. INTRODUCTION
Information Technology is playing an ever-increasing role in the provision of better and more efficient healthcare in the Western world. Many governments have recognised the benefits of utilising Information Technology in medicine, and, as such, have launched a succession of IT-related healthcare initiatives. A typical example is the United Kingdom's widely publicised commitment to electronic delivery of healthcare by 2008, as well as its NHS Cancer Plan [19], in which Information Technology features heavily.

SAC'04, March 14-17, 2004, Nicosia, Cyprus

One of the results of the technological developments—and associated deployment (successful or otherwise)—that have taken place in recent years is the highlighting of a number of major challenges. For example, the increasing range of imaging modalities, coupled with a very real fear of litigation, has resulted in clinicians having to cope with ever-increasing volumes of data, without necessarily receiving the benefit of appropriately focussed information. This problem is exacerbated by the lack of specialists in some areas.

The widespread appreciation of the well-documented potential benefits of Grid computing (see, for example, [12] and [2]) has increased the drive for the provision of distributed digital healthcare. The increasing importance of data-intensive applications to scientists from a wide variety of disciplines is driving the development of appropriate infrastructures, which—in turn—should drive some of the next major advances in the relevant application areas. Restricting ourselves solely to the area of healthcare, we see a number of potential advances. For example, the prospect of the development of teleradiology to facilitate the geographic separation of the skilled clinician from his/her less skilled colleague and that clinician's patient whilst improving diagnostic capability is one such potential development. The eDiaMoND project [6] has aims that are consistent with this.

The principal aim of the eDiaMoND project is to develop a prototype for a national database of digital mammograms to support the United Kingdom's breast imaging infrastructure. In this regard, eDiaMoND is similar to two other projects—the EU-funded Mammogrid project and the NDMA project [18]. As well as developing the a prototype infrastructure that is capable of supporting the breast imaging community within the United Kingdom, the project intends to develop a number of applications that take advantage of that infrastructure: comparison of images based on techniques described in [14] is one such application; others include a training application that has the potential to aid in the education of the next generation of radiologists, and a query interface to enable clinicians to conduct epidemiological studies, thereby taking advantage of the significant resource being developed. In the development of the system we aim to be consistent with other related initiatives, for example, the efforts of the DICOM and Health Level 7 (HL7) communities to bring together healthcare information system needs and medical imaging system needs [3]. We also aim to engage with—and take advantage of developments by—the wider Grid community (with the distributed query

processing efforts of [25] being a prime example).

eDiaMoND is funded by the UK eScience programme, and is an example of an emerging specialised sub-category of eScience projects—eHealth projects. One of the secondary aims of the eDiaMoND project is to deliver generic solutions to the problems faced during development and deployment—both technical and social. This paper is presented in that spirit—we feel that the approach taken to database design for the eDiaMoND project will be of interest not only to other mammography-related projects, but to a wide variety of applications in which DICOM is utilised.

The structure of this paper is as follows. In Section 2 we provide an overview of eDiaMoND, the project in which the work described in this paper has been undertaken. Then, in Section 3, we discuss the DICOM standard [17]. Next, in Section 4 we consider some of the other types of data stored within the eDiaMoND system. In Section 5 we describe our architecture for the transfer and storage of DICOM-related data (and, indeed, data from other sources). In Section 6 we describe the database schema. Finally, in Section 7, we discuss the contribution of this paper, and indicate some potential areas for future work.

2. THE CONTEXT

Breast cancer is one of the major health problems facing the Western world: it represents 19% of cancer deaths and 24% of all cancer cases in the European Union. Breast cancer is diagnosed in a total of 348 000 cases annually in the USA and EU [10]; it kills almost 115 000 women per year. Approximately one in eight women will develop breast cancer; one in 28 will die of it. The threat is negligible for women under 30, with the threat rising sharply until the age of 50, and continuing to rise (but less sharply) thereafter [31].

The earlier a tumour is detected the better the prognosis for the patient. A tumour that is detected when it is just 0.5cm has a favourable prognosis in about 99% of all cases. Few women can detect a tumour via self-examination if it is smaller than 1cm; typically tumours that are that large have been in the breast for up to 6–8 years. The five-year survival rate for localized breast cancer is 97%; this drops to 77% if the cancer has spread by the time of diagnosis [24]. This is the clear rationale for breast screening, before the introduction of which, 90% of breast tumours were found by self-examination [1].

Finland and the United Kingdom were the first two countries to develop national screening programmes, with many other countries following suit. (See [8] for an overview of national screening programmes.) In the United States, on the other hand, most mammograms are concerned with symptomatic patients [8].

Currently, the NHS Breast Screening Programme (BSP) (see [23] for an overview of the programme) invites women between the ages of 50 and 64 for breast screening every three years. If a mammogram displays any suspicious signs, the woman is invited back to an assessment clinic, where other views—using other imaging modalities—are taken. Currently, approximately 1.3 million women are screened annually in the UK. The Forrest report [11] was positive about the impact of the programme on the provision of healthcare in the UK:

> "by the year 2000 the screening programme is expected to prevent about 25% of deaths from breast cancer in the population of women invited for screening . . . On average each of the women in whom breast cancer is prevented will live about 20 years more. Thus by the year 2000 the screening programme is expected to result in about 25 000 extra years of life gained annually in the UK."

To date, the BSP has screened more than eleven million women, and has detected over 65 000 cancers. Recent research has demonstrated that the BSP is saving at least 300 lives per year [5]; it is anticipated that this figure will rise to 1 250 by 2010. However, despite the fact that screening has already produced encouraging results, there is much room for improvement. For example, it is estimated that 25% of cancers are missed at screening; the rate of such *interval cancers* has been greater than anticipated [30]. Such concerns inevitably give rise to resource-intensive suggestions for change. It has been demonstrated empirically that double reading greatly improves screening results [15]; indeed, it has been shown to half the number of cancers missed. It has been demonstrated that single screening plus the use of computer-aided diagnosis tools—image analysis algorithms that aim to detect microcalcifications and small tumours—also improves screening effectiveness.

The eDiaMoND project [6] aims to develop a Grid-enabled database of annotated mammograms, deployed at a number of sites in the United Kingdom. The eDiaMoND database has been designed with three key applications in mind: tools to support the diagnosis process; a training and education tool, and a query interface support of epidemiological studies. There are, of course, many other potential applications that could utilise such a resource. By taking three fairly diverse applications, it is intended that the design of the database will be sufficiently flexible so that the incorporation of future applications into the system will not be too onerous.

Specifically, the eDiaMoND project aims:

- to use the archive to evaluate innovative software that computes the quality of each mammogram as it is sent to the archive;
- to provide a huge teaching and training resource;
- to aid radiologists who are faced with difficult cases;
- to provide a huge resource for epidemiological studies; and
- to contribute, with regards to computer-aided detection, to the development of a technology that appears able to increase detection rates, and thus lives saved.

A project as ambitious as eDiaMoND could not be undertaken by any single organisation. There is a clear requirement for the developed system to satisfy the needs of the

end users if it is to be accepted: this project could not be undertaken without a detailed understanding of the procedures and needs of the radiologists, for whom the system is being built to support. In common with other complex UK e-Science projects, eDiaMoND is being undertaken by a project team possessing diverse and complementary skills; in particular, the team brings together a blend of expertise in clinical medicine, computer science and medical imaging. This paper reports solely on the work underpinning the database design.

The following have previously been identified as key issues to overcome with respect to the development medical image databases [32, 29, 28]:

- Large data sets: a decade ago, it was estimated that a major hospital typically generates in the order of one terabyte of digital imaging data annually [16]; this has increased significantly, and will continue to do so.
- Multimodality.
- Data heterogeneity.
- Structural and functional contexts.
- Imprecision.
- Temporal dimension.
- Infrastructure support.
- Security (techniques for ensuring security without impacting upon performance are very much in their infancy [27, 9]).
- Registration.

These issues—of course—are all of concern to us. In addition, however, the design of the eDiaMoND database has other, more Grid-related, considerations:

- Grid security: ensuring secure file transfer, and tackling the security issues involved in having patient records stored on-line, allowing access to authorised persons but also, potentially, the patients themselves at some time in the future is a key issue.
- Grid-enabling databases. (See [26] for an overview of issues in this area.)
- Data transfer: typically, each image is 30MB, or 120MB for a set of 4 images, which is the usual number for a complete case; issues here revolve around (loss-less) data compression and very rapid and secure file transfer.
- Ontologies and metadata: ontologies are being developed for description of patient and demographic data, together with descriptions of the image parameters and of features within images.
- Application-oriented issues: database design is being tailored to the obvious needs of rapid search and retrieval of images—clinicians have a justifiable demand for short response times when retrieving medical images [28].
- Data mining: some prototype data mining tools have already been developed within the context of a single database, with speed of access being a key factor.

Some of the above issues are peculiar to eDiaMoND, and our approaches to the tackling of them will be reported elsewhere. Rather than focus on these aspects in this paper—which would be of limited general interest—we choose instead to describe our approach to the capture and storage of DICOM files, which we feel is more relevant to the wider community.

3. DICOM: A BRIEF PRESENTATION

The Standard for Digital Imaging and Communications in Medicine (DICOM) is a widely used standard for the storage and transfer of medical images, which is based on two previous standards—ACS/NEMA 1.0 and 2.0, which were defined by the American College of Radiology (ACR) and the National Electronic Manufacturers Association (NEMA). What sets it apart from other image formats is that DICOM files contain both the image data and also image-related data. Image-related data can include, for example, data pertaining to how and why the image was taken, annotations, and patient information. This ability to capture non-image data is taken to its extreme when DICOM Structured Report Documents are deployed. In such circumstances, no image is included the file: it consists entirely of non-image data.

DICOM services can be divided into two groups. The first—*composite services*—are optimised for image interchange. The second—*normalised services*—are concerned with information management in a broader sense. In this paper, we concern ourselves with composite services.

The DICOM view of the entities related to medical images is given in Fig.1.

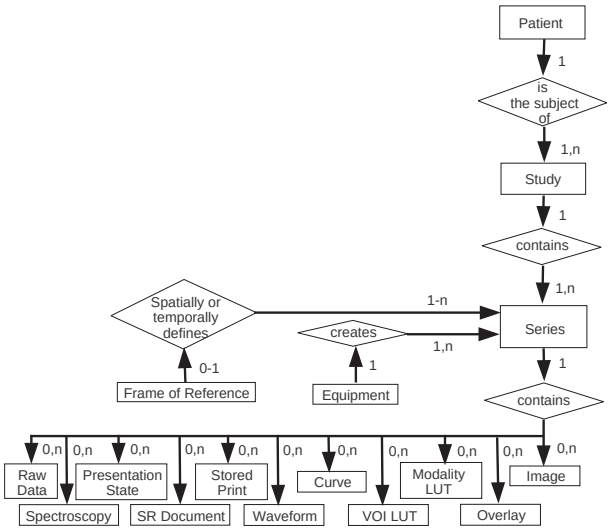

Figure 1: DICOM Entity Relationship Diagram

Here, a patient can be the subject of any number of studies. In the context of eDiaMoND, a study could represent the screening, then subsequent assessment of a patient; in

the course of a study one or more series of images may be taken. For example, in breast screening a single series of images would be taken using an X-Ray machine. If further assessment were to be deemed necessary then a series of images may be taken using a different X-Ray machine or possibly Ultrasound or Magnetic Resonance techniques. Typically, for each session—or for the use of a new piece of equipment—a new series is created. It is also possible to have a series of Structured Report (SR) Document files. SR Document files can refer to DICOM files (or simply the image part of such a file) belonging to other series without actually being part of that series; such referential power is available only to SR Document files.

Every DICOM file contains information pertaining to the patient, study, and series. It is common to store one image per DICOM file, so the patient, study and series data will be repeated in every file. This allows each image to be viewed in isolation without losing context, but involves storing the same data in many files and introduces the possibility of conflicting data in files that relate to the same patient.

Using the DICOM standard it is possible to exchange information relating to a wide variety of different image modalities. Within a single file—or as part of a single transfer—it is important that all parties involved understand what type of information is being exchanged. To facilitate this understanding, a reference is made to a Service Object Pair (SOP), which contains an Information Object Definition (IOD). It is the Information Object Definition that defines what type of object is being transferred.

Each Information Object Definition (IOD) contains a number of modules, with each module being related to an entity. Modules are not exclusive to IODs—most modules are utilised in multiple IODs—and it may be the case that they are related to different entities in different IODs.

Each module has a usage, represented by one of three letters: if a module has usage M, then it must be included; if it has usage C, then its inclusion will depend on an explicit condition; if it has usage U, then it is optional.

Each module is made up of a number of attributes. Each attribute has a name, a tag and a type. The tag is a pair of 16 bit numbers used to identify the attribute. The first number signifies the Tag Group that the attribute belongs to; the second number identifies the Tag Element. Together, the pair represents the specific attribute. Even for mandatory modules it is not necessary to use all the attributes. The type of the attribute determines if it must be present and whether it must contain any data: type 1 attributes must be present and contain data; type 2 attributes must be present but may contain no data; type 3 attributes are optional. Type 1C and 2C attributes also exist: such attributes are present only if certain conditions are met. It is possible that attributes may be present in more than one module; in this case the attribute may have more than one type. In these cases the most restrictive type is used.

In part 6 of the DICOM standard, each tag is given a value representation (VR) and a value multiplicity (VM). The value representation describes the exact binary representation of the attribute within a DICOM file. The value multiplicity gives the number of such attributes.

Having a value multiplicity of more than one allows simple lists of attributes; for more complicated repeated patterns of attributes, DICOM defines sequences. A particular attribute may be present in many different sequences and many times in a particular sequence. It is also possible for sequences to contain other sequences. While this would—at first sight—tend to suggest that an object-oriented approach to the design of the database might be appropriate, the drawbacks associated with object-oriented databases, such as, for example, the fact that query optimisation compromises encapsulation and the lack of a universally recognised data model [7] means that an object-relational approach is more appropriate. IBM DB2—which supports the eDiaMoND database—offers such capabilities.

4. DATA REQUIREMENTS

In this section we describe some of the types of data that we are concerned with.

The IODs currently supported by the eDiaMoND system are those pertaining to Secondary Capture Images, Digital Mammography X-Ray Images, and Mammography CAD Structured Report Documents. It is fully intended that future versions of the database will support other modalities, such as Magnetic Resonance Images, Ultrasound Multiframe Images, and Positron Emission Tomography Images.

Standard sets of patient data forms are gathered at the same time as image acquisition. Screening forms are stored as DICOM Structured Report Documents, which contain standard BiRADS classifications [4].

There are, of course, other types of information to be captured by the database. Data pertaining to audit trail and access control capabilities are of particular relevance, given the nature of eDiaMoND. In particular, there is a clear need to have a sufficiently flexible access control mechanism to allow individual hospitals or departments to enforce their own local policies. These issues are, however, outside the scope of this paper.

All updates to the database take the form of DICOM files inserted into the database. The non-image components of these files are stored in the database, together with an identifier pertaining to the user that inserted the DICOM file, as well as additional information, such as, for example, the time that the file was inserted. The database also keeps a record—stored in a log file—of all queries on the data.

5. ARCHITECTURE

Interactions with the eDiaMoND database occur via grid services. These grid services are realised using the Globus Toolkit 3.0 [13], which was used to produce Open Grid Services Infrastructure(OGSI)[22] compliant web services.

One such grid service is the query service, which allows users to query the database. Users are not permitted to query the database directly using SQL; instead, queries are sent in a pre-determined format as XML documents to the query service. Only allowing pre-determined queries affords nu-

merous advantages, with the key ones being the hiding of low-level details from the end-user and the opportunity for static application-level query optimisation.

The query results are also returned as XML documents, with the data being represented in XML WebRowSet format. Being an XML document it is relatively straightforward to transform the results into any format that is required using XSL transformations.

As can be seen in Fig.2, the query service communicates with the database through an OGSA-DAI Grid Data Service [20]; this is also an OGSI compliant grid service built using the Globus Toolkit.

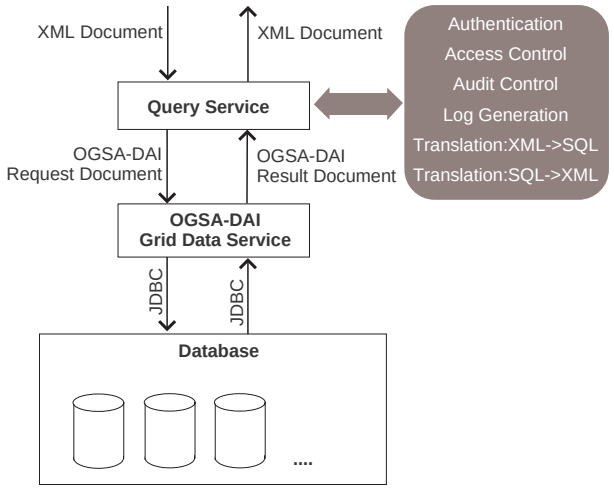

Figure 2: Grid Query Service Architecture

The query service can ascertain the access rights of an individual user by querying the database as a privileged user. Having determined the appropriate access permissions, the query is either rejected or translated into appropriate SQL. The SQL query is then placed inside an OGSA-DAI Request Document and passed to the OGSA-DAI Data Service, which queries the database using a JDBC connection.

The result of the query is then returned by the OGSA-DAI Data Service as an XML document. The query service extracts the WebRowSet data from the document and passes the results back to the user.

The query service also ensures that the access attempts are logged correctly by inserting the user details into the logs; this allows several users to be mapped to a single database account while maintaining traceability.

Queries are made against a single logical database—although, in reality—this will be either a federation of different physical databases or a group of replicated databases. There are several proposed strategies which are being investigated, including the use of an OGSA-DAI based distributed query mechanism OGSA-DQP[21]. Importantly, none of the proposed solutions will have an impact on the query service, which is agnostic to the underlying distribution mechanism being employed.

While it is desirable to be able to handle all types of DICOM files, it is not practicable—with the resources available to us—to create a normalised database that could fully represent an arbitrary file. This problem can be easily understood if the system was required to deal with a new version of DICOM, which—perhaps—introduced new data fields or new IODs. In effect, we are designing for forwards-compatibility. In addition, employing an unnormalised structure—with the potential performance benefits that may result—is inappropriate for an application that requires guaranteed consistency of data; data integrity is essential for an application such as eDiaMoND. Despite the fact that it is undoubtedly the case that performance is important to us, delivering correct data in a longer time is more preferable than delivering incorrect data quickly.

For these reasons we divide the eDiaMoND database logically into two parts, as illustrated in Fig.3.

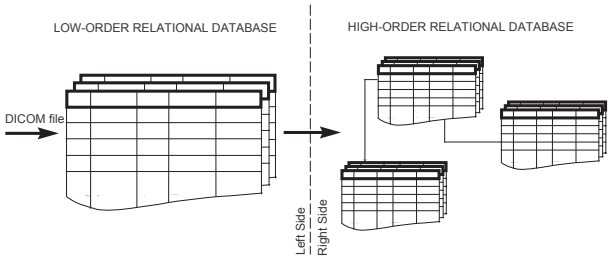

Figure 3: Database architecture

In the left-hand side—which we term *the repository*—the data is stored in a relatively unstructured fashion. In the right-hand side—which we term *the clinical information store*—data from specific IODs can be stored in a normalised fashion. When a DICOM file is inserted, it is parsed, with all non-image data being stored in the repository. Then, automatically, via automated constraint-enforcing procedures, the necessary data is inserted into the clinical information store. It is noted that in the repository, all the tags of the DICOM file are stored, including optional and private tags. In contrast, the clinical information store holds *only* the data that is currently useful for eDiaMoND applications. If any tags were to be deemed relevant in future, it would be trivial to recreate the appropriate part of the clinical information store using the existing data stored in the repository.

In this architecture there are some rules that have to be obeyed.

1. The repository allows only INSERTs, but not UPDATEs or DELETEs; this ensures that there is no potential for data loss.
2. The clinical information store allows only INSERTs and UPDATEs as a result of INSERTs to the repository; this ensures that the data in the clinical information store is consistent with that in the repository.

6. DATABASE DESIGN

The schema for the clinical information store is represented pictorially in Fig.4.

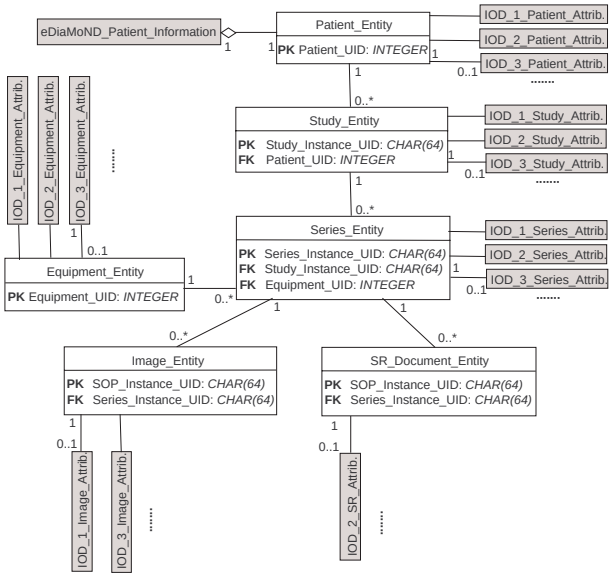

Figure 4: Schema to store DICOM data

This schema has been designed in a way that allows an arbitrary set of DICOM files to be stored in the database, while ensuring that the relationships between Patients, Studies, Series and Equipment is maintained.

The schema is logically divided into two parts: there is the common spine of DICOM entities and additional tables relating to specific IODs. In the diagram, three IODs are represented: *IOD_1* and *IOD_3* represent two different types of DICOM *image* IODs, while *IOD_2* represents a *Structured Report* IOD. While all three IODs have attributes related to Patient, Study, Series and Equipment, only the Image IODs have attributes related to the Image entity and only the Structured Report IOD has attributes related to the Document entity.

As was mentioned in Section 3, the modules of an IOD are all related to specific entities. Those modules contain attributes that may be present in more than one module, however the structure of DICOM is such that the same attribute will not be present in more than one entity. For this reason it makes more sense to group the attributes of an IOD by entity and not by module.

Although not shown in the diagram, the tables relating to each IOD are linked to the entity tables by a shared primary key, for example the *IOD_1_Image_Attrib* table has a primary key of *SOP_Instance_UID*, which has a foreign key constraint to the *SOP_Instance_UID* primary key of the *Image_Entity* table.

The *Patient_Entity* table is also linked to the *eDiaMoND_Patient_Information* table; this contains the additional patient data that was described previously, which is essential for epidemiology and integration with existing patient record systems.

The entity spine contains a bare minimum of data allowing efficient querying when using Unique Identifiers (UIDs). The *Patient_UID* and *Equipment_UID* are not part of the DICOM standard. The *Patient_UID* is represented as the *Patient_ID* attribute in the DICOM files, as an additional requirement on the users of the eDiaMoND system it is essential that all *Patient_ID*s are unique. The *Equipment_UID* is stored in the *Device Serial Number* attribute, and is also unique.

One of the most critical issues when developing a medical database is the provision of appropriate mechanisms for allowing updates and tracking changes. This importance is derived from the legal and ethical requirements to record all updates of patient and screening data.

In the eDiaMoND database, the deletion of previously captured DICOM files is forbidden. The principal reason for this is the necessity for keeping a history of previous data and cooperating—at the same time—with the existing health and legal regulations.

When an update of information is needed, e.g., a change of name or a change of address, this will take place as a two-phase operation. The first phase involves the insertion of a new DICOM file that contains the updated—or corrected—data. The second phase—which is triggered by the successful completion of the first phase—involves a copy and an update.

This second phase of the update mechanism is described by the following example.

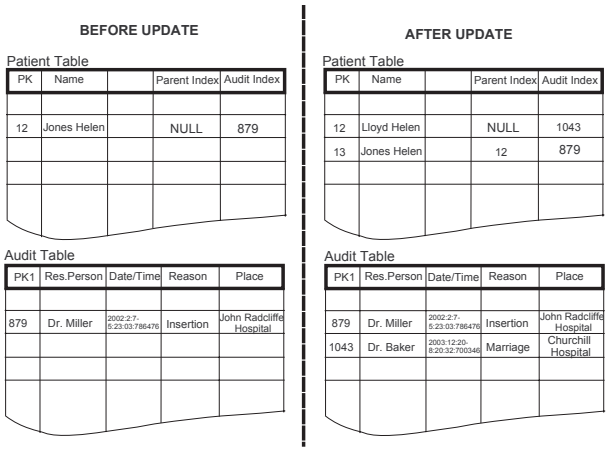

Figure 5: An example of the update mechanism

We consider a patient—Helen Jones—who has a unique ID: 12. Associated with this record is a related record—referenced by a foreign key called *Audit Index*—in another table, called *Audit Table*, which records the "who", the "when", and the "where" corresponding to the creation of a record in the *Patient Table*. If the patient were to get married, with a consequent change of surname, we would, of course, wish to update the patient's name, with all other pertinent information remaining the same. To achieve this, the patient record in *Patient Table* has to be duplicated, resulting in a new record, with a primary key of 13. The information

contained in the fields of the old record, having a primary key of 12, will remain the same, except the one that has to change, which in our case is the *Name* field, and also the field called *Audit Index*. The latter has to reference a new record in *Audit Table*, which will contain all the necessary information pertaining to the conditions under which the alteration has taken place. This new record is the one having primary key 1043 in the *Audit Table*.

In the above example we have used one *Audit Table* for a given table in the database. However, one *Audit Table* can be used to track the changes of many tables. The concept remains the same: the only modification is the addition of another field in the *Audit Table* containing a reference to the table that has been changed.

7. DISCUSSION

In this paper we have described an approach to the storage of DICOM files that has been taken in the development of a Grid-enabled medical image database. As DICOM is a widely recognised standard for the transfer and storage of medical images, the applicability of the approach is by no means restricted to the area of mammography. Furthermore, even though the eDiaMoND database is based on IBM's DB2 and Content Manager technologies, there is nothing vendor-specific about the *principles* that have been employed.

We feel that the design described has a number of benefits. First, the database has been designed with the storage of any DICOM file in mind: although we have restricted ourselves only to IODs of relevance to eDiaMoND, the database has been designed so that it can easily be adapted to handle other types of file or other IODs. This design decision—which has the fundamental aim of providing inter-operability, scalability and flexibility—was taken with the view of future-proofing the system: new epidemiological studies, new equipment and new reporting forms all have the potential to introduce new types of data into the database. The logical divide of the database into the clinical information store and repository means that one part can easily be restructured, e.g., the addition of further tables, without there being the potential for major disruption to the system as a whole. Second, the approach provides us with a full audit trail of inserted DICOM files; this audit mechanism allows both current and historical data to be queried. Finally, the flexibility of DICOM Structured Reports allows the potential for the use of BiRADS classifications.

With regards to future work, there are three immediate areas of opportunity. First, there is a clear need to design a flexible and fine-grained model of access control for the eDiaMoND database; again, this need will be replicated across similar projects. Second, the development of a Grid service to load and retrieve image files will be desirable. Similarly, closer integration of our research with that of the OGSA-DAI efforts is also high on the agenda. Finally, the development of semantic-level query optimisation techniques, i.e., enforcing domain-specific optimisation of a query, prior to any syntactic optimisation performed by the underlying DBMS, is a particular area of interest.

8. ACKNOWLEDGEMENTS

The authors wish to thank other members of the eDiaMoND consortium for their input into, and comments on, the work described in this paper. In particular, we would like to acknowledge the comments of Brian Collins, Sharon Lloyd, Jim Reimer, David Schottlander, and Paul Watson.

The authors wish to acknowledge the support provided by the funders of the eDiaMoND project: the UK Department of Trade and Industry, the EPSRC (ref no: GR/S20956/01), and IBM.

9. REFERENCES

[1] M. Baum and S. Denton. *Breast Cancer: The Facts*. Oxford Medical Publications, 1981.

[2] F. Berman, G. C. Fox, and A. J. G. Hey, editors. *Grid Computing: Making The Global Infrastructure a Reality*. Wiley, 2003.

[3] W. D. Bidgood Jr, L. Y. Korman, A. M. Golichowski, P. L. Hildebrand, A. R. Mori, B. Bray, N. J. G. Brown, K. A. Spackman, S. B. Dove, and K. Schoeffler. Medical data standards: controlled terminology for clinically-relevant indexing and selective retrieval of biomedical images. *International Journal on Digital Libraries*, 1(3):278–287, 1997.

[4] Breast Imaging Reporting and Data System: BI-RADS. www.acr.org/departments/stand_accred/birads.

[5] R. G. Blanks, S. M. Moss, C. McGahan, M. J. Quinn, and P. J. Babb. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990–8: comparison of observed with predicted mortality. *British Medical Journal*, 321:665–669, 2000.

[6] J. M. Brady, D. J. Gavaghan, A. C. Simpson, M. M. Parada, and R. P. Highnam. eDiaMoND: A grid-enabled federated database of annotated mammograms. In F. Berman, G. C. Fox, , and A. J. G. Hey, editors, *Grid Computing: Making the Global Infrastructure a Reality*, pages 923–943. Wiley Series, 2003.

[7] T. Connolly and C. Begg. *Database Systems: A Practical Approach to Design, Implementation, and Management*. Addison-Wesley, 2002.

[8] P. B. Dean. Overview of breast cancer screening. In M. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt, editors, *3rd International Workshop on Digital Mammography*, volume 1119 of *Excerpta Medica International Congress Series*, pages 19–26. Elsevier Science, 1996.

[9] M. A. Epstein, M. S. Pasieka, W. P. Lord, S. T. Wong, and N. Mankovich. Security for the digital information age of medicine: Issues, applications, and implementation. *Journal of Digital Imaging*, 11(1):33–44, 1998.

[10] J. Esteve, A. Kricker, J. Ferlay, and D. M. Parkin. Facts and figures of cancer in the European Community. Technical report, International Agency for Research on Cancer, Lyon, France, 1993.

[11] P. Forrest. Breast Cancer Screening. Report to the Health Ministers of England, Wales, Scotland and Northern Ireland. HMSO, London, UK, 1986.

[12] I. Foster and C. Kesselman, editors. *The GRID: Blueprint For A New Computing Infrastructure*. Morgan Kaufmann, 1999.

[13] Globus Toolkit. http://www-unix.globus.org/Globus, June 2003.

[14] R. P. Highnam and J. M. Brady. *Mammographic Image Analysis*. Kluwer Academic Publishers, 1999.

[15] K. Johnston and J. Brown. Two view mammography at incident screens: cost effectiveness analysis of policy options. *British Medical Journal*, 319:1097–1102, 1999.

[16] Y. Kim. Requirements for PACS workstation. In *Proceedings of the Second International Conference on Image Management and Communication in Patient Care, Kyoto*. IEEE Computer Society Press, 1991.

[17] National Electrical Manufacturers Association. *Digital Imaging and Communications in Medicine (DICOM)*, 2003.

[18] National Digital Mammography Archive. http://nscp01.physics.upenn.edu/ndma/.

[19] The NHS Cancer Plan. www.doh.gov.uk/cancer/cancerplan.htm.

[20] Open Grid Services Architecture - Data Access and Integration, OGSA-DAI. http://www.ogsa-dai.org, April 2003.

[21] Open Grid Services Architecture - Distributed Query Processor, OGSA-DQP. http://www.ogsa-dai.org/dqp/, August 2003.

[22] Open Grid Services Infrastructure (OGSI) Version 1.0. http://www-unix.globus.org/toolkit/draft-ggf-ogsi-gridservice-33_2003-06-27.pdf, June 2003.

[23] J. Patnick, editor. *NHS Breast Screening Programme Review 1996*. NHS Breast Screening Programme, 1996.

[24] V. L. Shavers and M. L. Brown. Racial and ethnic disparities in the receipt of cancer treatment. *Journal of the National Cancer Institute*, 94(5):334–357, 2002.

[25] J. Smith, A. Gounaris, P. Watson, N. W. Paton, A. A. A. Fernandes, and R. Sakellariou. Distributed query processing on the Grid. In M. Parashar, editor, *Proceedings of the 3rd International Workshop on Grid Computing (GRID 2002), Baltimore, MD, USA*, pages 279–290. Springer-Verlag Lecture Notes in Computer Science, volume 2536, 2002.

[26] P. Watson. Databases and The Grid. Technical Report CS-TR: 755, Department of Computing Science, University of Newcastle, 2002.

[27] S. T. C. Wong and H. Huang. Authenticity techniques for PACS images and records. In *SPIE 2435: Medical Imaging Proceedings—PACS Design and Evaluation, San Diego*, 1995.

[28] S. T. C. Wong and Kent Soo Hoo. Medical imagery. In V. Castelli and L. D. Bergman, editors, *Image Databases: Search and Retrieval of Digital Imagery*, pages 83–105. John Wiley & Sons, 2002.

[29] S. T. C. Wong, Kent Soo Hoo, R. C. Knowlton, R. A. Hawkins, K. D. Laxer, D. Tjandra, and M. Abundo. Issues and applications of networked medical imaging library. *International Journal of Digital Libraries*, 1(3):209–218, 1997.

[30] C. B. Woodman, A. G. Threlfall, C. R. M. Boggis, and P. Prior. Is the three year breast screening interval too long? Occurrence of interval cancers in NHS Breast Screening Programme's North West region. *British Medical Journal*, 310:224–226, 1995.

[31] J. L. Young, C. L. Percy, and A. J. Asire. Surveillance, epidemiology, end results: Incidence and mortality data 1973–1977, 1981.

[32] S. Zink and C. Jaffe. Medical image databases. *Invest. Radiol.*, 28(4):366–372, 1993.