Database Preservation Case Study: Review

Mette van Essen, Maurice de Rooij, Bill Roberts, Maurice van den Dobbelsteen
National Archives of the Netherlands
12 July 2011

Introduction

As part of the PLANETS project, the National Archives of the Netherlands carried out a case study on approaches for long term preservation of databases, published in May 2010 as PLANETS deliverable PA/6-D13.

Roughly one year later, the question of how best to maintain long-term access to information held in databases remains an important one for the National Archives of the Netherlands. As a first step towards further work in this area, we carried out a review of the PLANETS case study, triggered in part by participation in the Preservation of Complex Objects Symposium[1] in London in June 2011. This document is a copy of the original PLANETS case study with added commentary to document new developments in thinking, tools and technology.

At the end of this document, we have added two new sections, on the use of emulation for database preservation and the possible applications of data warehousing techniques in digital preservation. We have also added a new conclusion, briefly describing what we plan to do next in this area.

[1] http://www.openplanetsfoundation.org/events/2011-06-16-pocos-london-symposium

Project Number	IST-2006-033789
Project Title	Planets
Title of Deliverable	Case Study: Database Preservation at the National Archives of the Netherlands
Deliverable Number	D13
Contributing Sub-project and Work-package	PA/6
Deliverable Dissemination Level	Internal No restrictions
Deliverable Nature	Report
Contractual Delivery Date	31 March 2010 (M46)
Actual Delivery Date	27 May 2010
Author(s)	NANETH

Contributors

Person	Role	Partner	Contribution
Bill Roberts	PA/6	NANETH	Main Author
Maurice van den Dobbelsteen	PA/6	NANETH	Concept and planning
Andrew McHugh	PA/6	HATII	Reviewer
Amir Bernstein	PA/6	BAR	SIARD information and advice

Document Approval

Person	Role	Partner
Andrew McHugh	PA/6 WP Lead	HATII
Sara van Bussel	PA SP Lead	KB-NL

Distribution

Person	Role	Partner
Planets wiki		

Revision History

Issue	Author	Date	Description
1.0	Bill Roberts	27/05/2010	Document issue for approval

References

Ref.	Document	Date	Details and Version

1 Introduction

This case study describes investigations carried out by the National Archives of the Netherlands (NANETH) into how best to preserve digital records held in databases.

The purpose of the work is to consider the issues surrounding database preservation in an archival context, to help the organisation develop a preservation strategy for this type of digital object. To make the case study more concrete, we consider a specific example of a database that we want to preserve.

Our example database is one of the key outputs of the project DTNA (Digitale Toegangen Nationaal Archief), which is in the process of making all of the archive's *finding aids* available digitally, as a supplement to and ultimately a replacement for the traditional paper-based finding aids.

In this study we examine the archival requirements for preserving a database, review previous work in this area and consider the different requirements of different types of database. We review the SIARD database preservation suite produced by Swiss Federal Archives, developed partly within the PLANETS project, and examine its suitability for preserving our example database.

For further context, we also look at the software produced by the MIXED project ("Migration to Intermediate XML for Electronic Data"[2]), carried out by the Dutch project DANS (Data Archiving and Networked Services).

Based on this work, we present a simple gap analysis to identify what further work may be needed to establish a comprehensive approach to preservation of databases.

2 NANETH database of digital finding aids

A finding aid is a search tool which makes it easier to find material in a particular archive or part of an archive. The finding aids under consideration are for the paper collections of NANETH. Each archive within the collection has an inventory, giving an overall view of what it contains, for example "Documents received 1700-1780". Only an experienced searcher is likely to know what will be found based on this description. A finding aid gives a more useful and detailed specification of the contents of the archive, for example via an index of people's names or topics. Finding aids help to bring the archives closer to the general public, by making it easier to find information in complex archives.

In addition to directing a user to the correct original documents, finding aids often contain substantial and useful information in themselves, for example this database of emigrants from the Netherlands to Australia (http://www.nationaalarchief.nl/emigranten/). This database links the name of the migrant to a date of birth, date of departure from the Netherlands and date of arrival in Australia and the ship the migrant travelled on: as well as identifying the location of the original record, in this case the emigration card.

Another example of a complex finding aid can be found here: http://www.en.afscheidvanindie.nl/ providing information on the Dutch withdrawal from the former Netherlands East Indies, drawn from a large number of sources in the Netherlands and in the archives of several other countries. It provides a searchable index into the original source material, listing the types of information that is available and where it can be found. Some of the original documents have been photographed and are made available online, for example http://beeldbank.nationaalarchief.nl/viewer/schaduwarchief/2.22.21/376/0003

[2] http://mixed.dans.knaw.nl

The DTNA project will be complete in the middle of 2010 and at that point we must consider how the valuable information produced by this project can be preserved for the long term.

In the past, the Archive's finding aids have been in a variety of forms, from paper to spreadsheets to individual databases. The DTNA project has taken these finding aids, processed them into a standard format, cleaned and checked the data and added useful metadata. The processed finding aids are then stored in a single MySQL database. When complete, there will be approximately 200 finding aids in the system, with a total of over 2.5 million individual records.

To provide public access to the finding aids, individual finding aids are packaged up and transferred from the DTNA database to the NA4All system (NANETH's public web portal). The transfer packages are made up of metadata in a standard XML format, together with CSV files for the finding aid content. These packages can be automatically processed by NA4All to create a web interface to the finding aid.

2011 Review Comments

The DTNA system has been in active use since the case study was published and as of June 2011, DTNA holds 127 finding aids with a total of 3.15 million records.

The DTNA project (migrating 5000 pre-existing finding aids into a single database) has been a good start in developing a preservation strategy for this information, because it is now in a standard format, cleaned up and checked. However, standardising and normalising alone does not assure preservation. For long term access further preservation strategies must be tested.

.

3 Review of Digital Preservation Testbed recommendations

The Digital Preservation Testbed project, carried out by NANETH between 2000 and 2003, published recommendations on the preservation of databases[3]. It identified the main components of a database system as:

- Database content
- The Database Management System (DBMS)
- Software applications used to access the content of the database for a given purpose

The principal concern of a central government archive such as NANETH is to preserve the records of government. For that reason, it is important to consider how the components of a database system relate to archival records and to business processes.

Which aspect of the database system corresponds to the archival record? This may differ from one case to another. Possibilities include:

- The complete database system: content, DBMS and application(s)
- The database content alone
- A single row of data in a table in the database
- Data distributed over multiple tables
- Information taken from the database and displayed on screen by an application

Furthermore, should we be interested in the final state of database once it is no longer in use, or the state of database at a given time, or the history of changes to the database? Again, this may

[3] http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-databases-en.pdf

differ according to how the database relates to the business processes of the organisation where it was used.

The Digital Preservation Testbed project identified the following authenticity requirements for databases:

- Content: this is primarily the contents of the database tables, but may also include how the data was grouped and presented on screen. Note also that databases can be used to store complex objects such as images or documents, therefore it may be necessary to treat a database as a compound object with different preservation strategies for objects of different types held within the overall database structure.
- Context: metadata should be retained about the organisation that used the database, how the data was used in business processes and how information from the database was used by applications. This latter requirement could be satisfied in some cases by storing the queries used to extract data from the database for a particular purpose.
- Structure: of the database – how the data is organised into tables and interlinked.
- Appearance: how the data was presented on-screen (or on paper reports) by the accessing application.
- Behaviour: of the accessing application may be important in some cases.

The Testbed recommendations for a preservation strategy noted that emulation was of interest because of its potential to maintain access to the original end user application, but at the time of the project, no suitable emulators were available to test this approach.

Therefore the recommendation was that the data in the database should be converted to XML or another well documented standard format and that metadata and documentation should be provided to explain how the database was used. The conversion of the data should be carried out as early as soon as possible after it has been decided that the data should be preserved, as most tools for database migration require a working instance of the original DBMS.

Since 2003 when these recommendations were published, understanding of emulation has developed and specific emulation tools are available that would be capable of reproducing personal or desktop databases and applications, for example those based around Microsoft Access. For more complex databases, particular those with a client-server architecture, it may be necessary to emulate two or more machines simultaneously (eg one for the server, one for the client, possibly also an intermediate application server) including the network communication between those machines. This is feasible in principle but is beyond anything that has been tested in the PLANETS project.

An early version of the Swiss database preservation tool, SIARD, was tested in the Digital Preservation Testbed project and it was evaluated at the time to be a good approach for preserving the content of a relational database.

The main uncertainty of the Digital Preservation Testbed database research and the associated recommendations was identifying precisely what constitutes the archival record when using databases in the business processes of an organisation. Without a clear identification of that, it is difficult to make reliable recommendations on how to preserve the records, and as described above, how the archival record corresponds to the technology used to support the business process is a complex question.

Answering these questions was not part of the scope of the PLANETS project and to the best of our knowledge this has not been investigated in detail by other researchers since the Testbed recommendations were published. Therefore it remains an area requiring further investigation. This case study helps to define the problem and review possible approaches, but more detailed work will be required.

2011 Review Comments

We still have no definitive answer on the question what constitutes the archival record. As discussed in the next chapter of the case study, we expect to encounter different situations with different requirements. Whilst the use of databases is widespread in many of the important

business processes of the government ministries served by the archive, it remains unusual for the archive to receive databases into its custody.

An important conclusion of this review is that we should actively discuss database archiving requirements with ministries, to find out more accurately what we will have to deal with in the future.

4 Archival requirements

We need to consider the requirements for authentic preservation of records associated with databases. We can identify the following three scenarios:

1. Preserving records of background data used to inform or justify a significant decision
2. Preserving records of important events or transactions that were stored in a database.
3. Preserving structured information of historical interest.

For each of these we must consider which of the main components of the database system must be preserved: the content of the database, the database management system and the applications that used the data. Or are the records of interest in the form of reports produced from the database that can be preserved separately from (or instead of) the database itself?

Scenario 1: Preserving data used for decision making
In this case it is required to record the information used to inform or justify a significant decision. An example might be an agency responsible for environmental protection deciding whether to renew a permit for a company to release waste water into a river. The decision might depend on the history of water quality measurements in the river, information on the plant and animal species that live in the river and monitoring of the company's past compliance with the terms of the permit. All of this kind of information would typically be held in a database.

A typical process would be to extract relevant information from the database using one or more queries, then to present that information in a report. Therefore to preserve this information, one could either preserve the entire database and the queries used to extract the information for a particular decision; or to preserve the extracted information as it was used in the report. This latter process is likely to be simpler and to provide a clearer link between the decision and the information that was used. Therefore in this case the preservation requirement becomes focused around documents rather than the database itself. Only those data directly used in the decision are part of this record. (It may be decided that the entire monitoring database is of interest in itself and this situation is discussed further in Scenario 3).

So in this type of scenario there is (in most cases) no need to preserve the application that was used to extract the information from the database, only the results of that extraction process.

2011 Review Comments

Using this approach, we must be cautious of preserved data containing links back to the database contents, in case the linked content could change after the report was created.

Scenario 2: Preserving records of events or transactions
To continue the example of an agency issuing permits, the agency must also keep a record of all applications for permits and whether the permit was granted or refused, by whom, on what date and so on. Again in most cases this information will be stored in a database. Assuming the same database is used year after year to hold this information, the archiving process may involve taking the older records out of the 'live' system and transferring them to the archive. This could be achieved in a number of different ways, but one example could be to keep (say) the last 10 years

of records in the live system and at the end of each year to transfer records older than 10 years to the archive, so at the end or 2010 the records for 2000 would be archived. The archiving and preservation task would then be to ensure that the database of records for the year 2000 remains accessible.

In this case, the preservation requirements are to preserve the contents and structure of the database, as well as contextual information about how the database was used in practice. It would not seem necessary to preserve the application that was used to work with the database in order to preserve this type of information. This scenario is therefore well suited to tools like SIARD.

2011 Review Comments

We note that to make use of data preserved by SIARD, it is necessary to read the SIARD data back into a relational database system. SIARD currently supports both extracting and reloading data with three RDBMS: Oracle, MS SQL Server and MS Access.

Scenario 3: Preserving information of historical interest

There are also examples of databases that are not directly related to the business of government, but are nonetheless of historical interest and worthy of being retained by the archive. In this case it is most likely that a database will be preserved in its entirety as a single digital object, rather than taking year by year slices of it as in the previous example. The DTNA database of finding aids most closely fits this scenario. In fact the DTNA database is a combined version of information from many individual spreadsheets and databases. The options for preserving the DTNA database are discussed in more detail in the next section.

In the case of the DTNA database, the important information to preserve is held in the contents and structure of the database. However, there may be other types of database where it is also of interest to preserve the application that was used to access and present the data if this is viewed to be of cultural/historical interest. In that case an emulation approach may be the only option, but such examples are probably rare.

2011 Review Comments

For this kind of database a combination of migration and emulation may be a good strategy. We can make a distinction between:

1. Extracting the records to preserve (only) the contents of the database
2. Preserving the contents and the structure of the database – using tools like SIARD or MIXED to create an intermediate format suitable for long-term preservation.
3. Preserving the database as a single object, using emulation to maintain the capability to run the application which can understand the database object.

5 Preservation requirements for the DTNA database

The preceding sections have explained the issues that must be considered when preserving a database. Here, we consider how this applies to our example database of finding aids.

As explained previously, all the finding aids are held in a single MySQL database. Because this incorporates data from approximately 200 separate original finding aids, some of which were databases in their own right, the structure of the database is relatively complex. The DTNA database allows the provenance and processing history of each individual entry to be traced.

There are three potential forms of the finding aids that could be preserved:

- The format prior to processing by DTNA.
- The MySQL database of all finding aids
- The XML and CSV format exports of individual finding aids from the DTNA system to the public portal NA4All.

The first of these can be discounted because it would fail to take advantage of the thorough data cleaning and checking work that has been carried out by the DTNA project. Moreover, the preservation task would be made more complex because it would need to deal with a large number of different formats of information.

The third option is feasible, but would not capture all of the information that is in the MySQL database, notably some of the provenance information.

The finding aids sometimes include external references, for example pointers to images in NANETH's "Beeldbank" system (an online database of images). Therefore a complete approach to preservation of the finding aids would require that the images and system of identifiers in the Beeldbank are also preserved.

Therefore our favoured approach is to preserve the full MySQL database as a single entity. After the DTNA project finishes in mid 2010, it is anticipated that this approach will be tested, making use of SIARD as the main preservation tool. If successful, the transformed data will be ingested into NANETH's Digital Depot system for long term storage and managed preservation.

SIARD is well suited to this task as it is the content and structure of the database that is important to preserve. In this case the way that information is accessed is not important for the long term value of the record.

2011 Review Comments

Our original plan of preserving the database as a single entity making use of SIARD and ingesting this into NANETH's Digital Depot System has not yet been tested, mainly because SIARD does not yet support importing from MySQL.

The primary purpose of the DTNA project was to collect finding aids and make them accessible through a convenient, standardized system. However a useful side-effect is that the finding aids have been made much more suitable for long-term preservation as they have been reviewed, cleaned-up and organized into a single well-structured system.

The processing of the finding aids has taught us many useful lessons relevant to preservation of databases and other structured data. It revealed that there is no such thing as a standard way to import data. Most of the 3.1 million records needed some kind of human intervention during the import process. The data of the DTNA project was imported using a variety of different methods such as direct database connections and exporting data as CSV from the source.

It is also clear from the DTNA experience that well-designed, carefully structured data sources, whether they are simple databases or spreadsheets or text files, are much easier to understand and process. However it is often the case that we end up trying to understand and work with poorly designed or poorly maintained data. It would be highly beneficial for long-term preservation if the application of good practices in database design could be encouraged.

When the DTNA project finished it delivered a clear set of guidelines regarding cleaning and normalizing of data that can come from any source: CSV, plain text, MS Access, DBF, MSSQL, etc. These normalization principles could be very useful when thinking about the next generation database preservation tool or planning preservation of databases.

6 Database preservation tools and methods

6.1 SIARD

Background

SIARD (Software Independent Archiving of Databases) was initially developed by the Swiss Federal Archives as part of its ARELDA programme, then was further developed and refined within the PLANETS project.

SIARD is in active use by the Swiss Federal Archives (SFA), to handle databases primarily relating directly to government business and supporting information for government. The most common database engine used by the Swiss government organisations is Oracle: but the SFA also must deal with Microsoft Access and Microsoft SQL Server databases.

The content of these databases is varied, with the most common content type being personal information such as names, addresses, police data etc. More complex data such as scientific data is rare. Some agencies handle some statistical data.

Specific examples include:

- LADIS, the databases of the Swiss Air Force, containing aerial images
- IPAS, the police database
- ZAR, the central database on foreigners in Switzerland
- The database of the Migration Authorities
- Agricultural information such as The Fruits and Berries Database of Switzerland

Both live and decommissioned databases must be archived. For those databases which are regularly updated, for example the database of the Customs Administration, the SFA uses SIARD to take regular snapshots. They also deal with dumps of complete databases that are no longer in use. Another scenario is to archive dumps of 'archiving databases', which are part of an active central database, or are created from an active central database, to hold those parts of the data that are no longer in active use. After archiving, those tables are subsequently deleted from the original central database.

In response to requests from the SIARD user community, work is in progress to add MySQL to the list of database types that can be preserved with SIARD.

2011 Review Comments

The latest release of SIARD was in April 2011 (Version 1.23). Support for MySQL has not yet been implemented.

Using SIARD

As part of this study, SIARD was downloaded and installed on a Windows PC, then tested with a Microsoft Access database. It is based on Java, so the software runs on any computer with an appropriate version of Java installed. A comprehensive user guide is available.

In order to connect to the Access database, some preparation was required: it was necessary to modify the access permissions to three 'system tables' inside the database and then to create an ODBC data source in Windows. The necessary steps are described in the SIARD documentation.

With those preparations in place, it was straightforward to use SIARD to convert the data into SIARD's standardised format. The contents of the tables are represented as XML files, the structure of the database is represented in SQL3 format, and 'Large Objects' (such as files) held within the database are stored as separate files. All the files created by SIARD are held inside a single ZIP64 container, so that the final output of the process is a single file with extension '.siard'.

For access to the data, SIARD also provides tools to transfer the contents of the .siard file back into a running database, with the choice of MS Access, Oracle or MS SQL Server. In this study, the SIARD data was transferred back into a new MS Access database and compared with the original. All important aspects of the database contents were preserved during this round-trip process. The original Access database also included a number of forms and reports for entering and presenting data. SIARD does not attempt to preserve these, as they are closely tied to the Access execution environment and cannot be represented in the standard storage formats used. Therefore these forms and reports were no longer present when the data was transferred back to a new Access database. Whether this type of information is important must be assessed on a case by case basis, as discussed in the earlier section on Archival Requirements.

SIARD connects to a running copy of the database (or to a Microsoft Access file via an ODBC connection) and extracts information using SQL, the standard "Structured Query Language" for relational databases. It therefore requires a running copy of the database at the point that SIARD is used. This means that it is well-suited to most database preservation situations where the data has been actively managed, but may not be applicable to historical database files where the original database engine is obsolete or no longer available.

2011 Review Comments

In 2010 the Swiss government adopted SIARD as its official format for transfer of relational databases to the Swiss Federal Archives.

6.2 MIXED

The MIXED framework ("Migration to Intermediate XML for Electronic Data") is 'a tool to extract tabular data from legacy data objects and convert them to a preservation format'. It is currently in development by DANS, ("Data Archiving and Network Services"), an institute within the Royal Netherlands Academy of Arts and Sciences (KNAW).

It is still under development and a preview version was provided to NANETH for the purposes of this case study. It comprises a set of command line tools for converting from database and spreadsheet formats to a format suitable for preservation and currently runs on Linux (or on Windows via Cygwin). The supported original data formats at the time of writing are Microsoft Access, dBase and Dataperfect.

The preservation format for relational databases consists of provenance metadata in XML format and the content of the database represented in (a simplified form of) the SIARD format, which is itself a combination of XML and SQL.

When preserving data from spreadsheets, MIXED makes use of the Open Document Format, an XML based format originally developed for the OpenOffice suite, but now an ISO standard.

The currently available tools in MIXED operate directly on the original database binary files, rather than via an interface such as JDBC or ODBC (although it would be straightforward to add such JDBC tools to the MIXED framework). This is of interest as it enables data to be extracted from such database files without needing a running copy of the original database management system.

2011 Review Comments

According to the MIXED website[4] in June 2011, the supported formats are: Data Perfect, Access 2000 and 2002, dBaseIII and IV and Excel 2003.

The MIXED project ended in 2010. DANS continues to work on the outcomes of the project so that these can be used not only by DANS but also by the community.

7 Gap Analysis

The preservation approach taken by SIARD is applicable to a large class of databases of interest to NANETH, the software works well and benefits from ongoing development and support by the Swiss Federal Archive. It works with three of the most commonly used database management systems (Oracle, MS SQL Server and MS Access) and an extension of the system is planned to add another (MySQL).

Therefore SIARD can make a significant contribution to the database preservation requirements of NANETH, but there are still issues to consider, most importantly:

- A more detailed analysis is required of how databases used in government relate to business processes and which aspects and which representations of the information constitute records which must be archived and hence preserved
- How to deal with databases in now obsolete formats, where the original database application can no longer be run (the MIXED tools may be able to help here, as might running original software in an emulated environment).
- How to identify in which cases it is important to preserve the application used to access the data and how to solve that problem.

2011 Review Comments

Databases may not simply hold 'data', but in many cases may hold other digital objects in a range of formats. Such complex compound digital objects may require a combination of preservation strategies.

Identification tools for complex data objects need further research: not only to identify the relevant database application and format, but to identify the formats of objects potentially held within the database.

8 Conclusions

The principal conclusions of our research are as follows:

- NANETH will be responsible for archiving and preservation of databases and needs to develop procedures for carrying this out.
- Tools already exist to transform information in databases into a format suitable for long-term preservation. Of these SIARD is currently the most appropriate and best developed option and would be a good solution for many databases. The tools being developed by the MIXED project are also of interest and should extend the range of database types that can be handled.
- Further investigation is needed on the fundamental issue of which aspects of databases and their use in government organisations will produce records which must be archived.

[4] https://sites.google.com/a/datanetworkservice.nl/mixed/

- In some cases it may be important to preserve the ability to run applications which interface with databases and more research is required on how to achieve this.

9 Acknowledgements

Thanks are due to Amir Bernstein of the Swiss Federal Archives for assistance with SIARD and information on how it is used in practice, Jan van Mansum of DANS for providing information on MIXED and Maurice de Rooij of the DTNA project at NANETH for advice on DTNA and assistance in assessing SIARD

2011 Review: Database Preservation and Emulation

The case study shows that in many cases, migration to a suitable 'preservation format' is a good preservation strategy for databases. However, we have identified clear cases where "simple" preservation does not suffice or does not work.

These cases are:

- Where there is a significant interactive element that is important from an archival perspective
- Where there is a need to preserve the original "look and feel" of the object
- non-standard material with interactivity: complete databases, CAD/GIS information, wiki's etc.
- databases from personal or other non-governmental archives, where the system has not been actively managed and relies on software which is no longer available.

In these cases the use of emulation to maintain the ability to run the original software application may be the best approach. Emulation may also be a useful component of a migration strategy, enabling obsolete software to be executed to convert digital objects from inaccessible to accessible formats.

Since the end of the PLANETS project, another EU research project, KEEP (Keeping Emulation Environments Portable) has continued to investigate and develop tools to support the use of emulation in preservation. KEEP has recently released the first version of its 'emulation framework'[5] which simplifies the process of setting up emulators and associated operating systems and software applications.

To our knowledge, emulation as a preservation strategy for databases has not yet been thoroughly tested in practice and further investigation would be very valuable.

2011 Review: Data Warehousing and Digital Preservation

The June 2011 Preservation of Complex Objects Symposium was followed with an associated short workshop on Data Warehousing in a digital preservation context, organised by the University of Portsmouth.

Data warehousing is an established technique for managing large complex datasets in a way optimised for effective querying (or 'data mining'). Janet Delve and colleagues in Humanities Computing at the University of Portsmouth are researching ways in which the techniques of data warehousing could be applied to the problems of digital preservation.

One such application is to use a data warehouse for the metadata for digital objects, essentially representation information. More complex objects, for example multimedia or 3D model data, typically require more description, hence more metadata to manage. The relationships between digital objects and the stack of software and hardware components needed to access them can be complex and data warehousing techniques have been proposed as a way to handle this.

[5] http://sourceforge.net/projects/emuframework/

2011 Review: Conclusions

Through the Digital Preservation Testbed project, PLANETS, DTNA and other activities, the National Archives of the Netherlands has investigated and gained experience of the issues around long-term preservation of databases. However, we have still relatively little experience of doing this in an operational way for databases used in the process of government.

To advance our knowledge and capabilities, we think the greatest value can be gained through a focus on finding practical solutions to 'real world' problems. To this end we aim to do the following:

- engage actively with ministries to understand the problem better and to help records managers in the ministries identify which databases will need archiving in future
- carry out experiments to evaluate preservation options for real databases from the organizations that the archive supports.