

DLM forum 2005 conference: preservation and digital archives, 6 October 2005, 16.30

'Rotterdam Municipal Archives: a digital repository, learning by experience with open standards and open source'

Klaartje Pompe, Rotterdam Municipal Archives

This paper is about a project in Rotterdam that aims to build a digital repository. It was started in 2004 by Rotterdam Municipal Archives and the Archival School (Archiefschool), the Amsterdam based institute for archival education and research. It is a remarkable project because:

- It is the first initiative for an archival institution in the Netherlands to test the existing theory, standards and international examples by practical experience;
- It started as a low cost project and although costs in 2005 have gone up, it still is;
- The starting point of the project is pragmatic. We use existing theory and it is being done by staff of our own organisation, Rotterdam Municipal Archives. The Archival School brings in knowledge and advice;
- The results of the project are to be shared with the Dutch archival community.

The subject of this pragmatic Rotterdam project is of interest to other organisations willing to start a digital repository because it shows a possible strategy how to start designing and developing. In Rotterdam there were no extensive fore studies, no deterring costs, at the start only a director and staff committed to the idea.

Of course there are pitfalls in this approach as well:

- The strategy of 'trial and error' with open source products should not continue indefinitely. Too much failures don't motivate;
- The dependency of the commitment of project members;
- Project members can follow their own course and insight;
- Cyclic developing has the danger of too much repeating without validating.

My presentation deals with:

- An outline of the problem;
- International and Dutch initiatives;
- A sketch of Rotterdam Municipal archives, its starting points and characteristics;
- The project in 2004 and 2005. What are the results and setbacks of this low cost, open source approach where learning stands central for all those involved?

Problem

The challenge for every archival institution now or in the near future is long-term preservation. One solution is the realisation of a digital repository for digital information objects, such as archives, digital publications, photographs, sound and moving images.

Rotterdam Municipal Archives has the ambition to become a trusted digital repository for digital objects as well as for analogue; in our mission statement should be guaranteed the authentic, reliable, long term accessibility of digital objects in our administration.

The challenge goes beyond the technique; developing and building a digital repository concerns the organisation, its strategy, staff, resources, procedures and methods.

International and Dutch initiatives

In the Netherlands there is no practical experience in the realization of a digital repository by an archival institution. The Dutch Royal Library (Koninklijke Bibliotheek) has together with IBM developed a digital repository for e-publications in the Netherlands. The National Archives have

written down functional specifications for a digital repository in 2000, not having resulted yet in the developing of a digital repository.

The aspects of cost and uncertainty (where to start, what are the consequences for organisation and staff members) seem to constitute barriers for most archival institutions.

All over the world a lot of research has been done about requirements, methods and techniques for the long term preservation of digital objects. Rotterdam Municipal Archives uses existing research and theory, there is no question of redoing work.

We are grateful to use the results of the Dutch Testbed project, Erpanet publications and of course the Interpares I and RLG project. These forms of research are very fundamental.

The published experience and results of The National Archives of the United Kingdom, the National Archives of Australia (NAA) and City Archive of Antwerp help us as well.

In the mean time, worldwide archival institutions develop and build digital repositories. When looking at these impressing initiatives, they give us input for our work in Rotterdam (strategy, applied standards, information) but when looking at the costs of these projects we are impressed as well. I mention the repositories of the National Archives of the United Kingdom, National Archive of Australia and our Dutch Royal Library (KB). Of current interest (8 september 2005) is the contract given to build US Electronic Records Archives (ERA) of the National Archives and Records Administration. Costs: 308 million dollar for a 6 year program. Digital preservation is an urgent problem and very complicated as well, but if it can only be solved with millions of euros or dollars, we in Rotterdam or other smaller parties compared to the above mentioned organisations face a huge problem. A problem for the whole administration and cultural heritage sector.

Rotterdam Municipal Archives

Rotterdam Municipal Archives is the second largest Municipal Archive in the Netherlands. We aim to make sure that archival objects in a wide range of media, from medieval manuscripts to digital records, remain authentic and trustworthy for present and future use.

Rotterdam Municipal Archives has a staff of 114 people. 2 repositories house 17 km¹ of archives, 90.000 publications, 325.000 photographs, 600.000 negatives, and thousands of films, videotapes, audio cassettes and placards. Staff has the know how and procedures in house to administrate, preserve and deliver to the public these physical objects.

In 2003 the management and staff recognised that in increasing quantity digital objects confronted the institution to be ingested in a repository and recognised the urgency to act. These digital objects in 2003 were mostly documentation, e-publications, digital photographs. Important to save were the numerous Rotterdam websites, an informative sources about the town and its citizens.

The confrontation with archiving digital objects raises questions about:

- Organisation: what are the consequences for the organisation, its strategy and staff;
- Staff: what knowledge to acquire;
- Technique: are we going to do the storage custodial (in house) or non custodial. What storage capacity will we need. What software is available and suitable for performing functions in a digital archiving system;
- Procedures: with what procedures to preserve digital records, created by the local government and private creators;
- Resources: our organisation has no extra resources, but digital preservation would eventually require extra budget in the end;
- User needs and expectations: how are we going to present our online information;
- What to do with our digitised objects; scans of local newspaper, appendixes of marriage acts, visual material. Remarkable is that until 2005 we had no digital records to store; digital publications and collected photographs were the first objects that needed to be ingested. And of course we have large quantities (700 Gb) of digitised material in

house. These originally analogue, non digital born objects need proper custody as being an asset.

Project in 2004 and 2005: just do it!

To answer these questions the Rotterdam Municipal Archives started in 2004 together with the Archive School a project E-depot to realize a prototype of a digital repository. A repository on the basis of open standards and open source software and the building of the digital repository at low cost. The members of the project team came from the organisation itself. Theme for 2004 was 'E-repository: just do it!'

Quite small, with a motivated but still inexperienced team, one server and open source software we started the project.

Learning stands central in the project. Staff of the organisation should acquire the same knowledge about and professional attitude towards digital objects as they have about physical objects.

Starting points

Starting points in the project were:

- No redoing work; using work already done by others;
- Using open standards: OAIS reference model, XML. In the nearby future we will probably work with EAD and ISAD (G) and ISAAR (CPF);
- Starting with open source software (DSpace for metadata and storage, i-Tor for accession). The reason why open source software has been chosen in the project is threefold:
 - It is non proprietary;
 - The community of users is supposed to assist and work together in developing;
 - It is more durable than proprietary
- Choice for migration as strategy, in most cases migration to XML.

In 2004 the prototype of the digital repository worked with DSpace, open source software developed by Massachusetts Institute of Technology en HP-labs. DSpace is developed as e-repository software for university libraries; it has served in our approach of 'trial and error' with software in 2004 and 2005. During 2005 we were confronted with the limitations of DSpace as software for a repository of an archival institution. We have chosen to develop software now modular and for the processes (Quarantaine, Preservation) not covered by DSpace. In the end of the project year we will have to decide whether to continue with open source software.

For the interface and accession we use open source software, that was developed by a Dutch research institute. The international open source component in software was further enlarged with XENA, software developed by the National Archives of Australia, to convert and deconvert documents in a durable format (XML).

2005: designing, developing and changing

The small project of 2004 has grown, widened and deepened in 2005; 25% of our staff members are involved in the project. The project team has divided in ten little workgroups who research and experience the archiving of websites, databases, large files, multimedia, the necessary hard- and software, the accessibility of digital objects from the repository. All together small but important building stones for a complete digital repository.

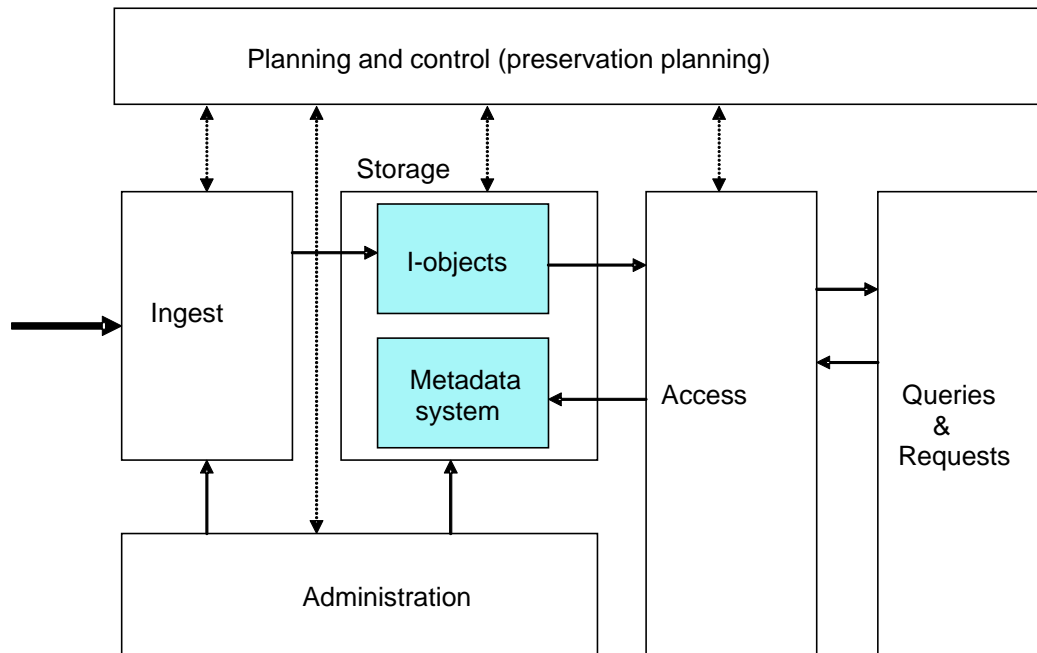
Basis for our design is our Business model, related to the OAIS-model. Herein we distinguish the following four operational work processes:

- Ingest and storage
- Administration (maintenance of conservation, migration, description)
- Access

- Queries and requests

Planning and control (preservation planning) directs all these processes.

Figure 1



Hard- and software

In 2005 we develop the e-repository approaching the overall hardware architecture of the National Archives of Australia (NAA). Like NAA we have separate components in the repository:

- Quarantine: digital objects enter the Quarantine by dvd or LUN. Checksums are verified and virus checks are done. After 4 weeks another verifying of checksums and virus check;
- Preservation: preservation processes are applied to the objects to ensure that they remain accessible over time. This means conversion to a durable format by the Australian open source software application XENA. Describing metadata will be added;
- Repository: here is storage for the long term of the original bitstream and the XML version of the object;
- Accession: in the end we will have one interface for accession of our digital born and digitised objects.

Costs 2004-2006

At the start of 2004 we had budget for courses, software adaptation and external expertise. It was possible to start the a digital repository project with:

- Knowledge and advice from the Archive School;
- Staff of the own organisation, committed and willing to learn;
- Existing theory and use of open standards;
- Minimal hardware; one server;
- Open source products.

At the end of the project year 2004 we bought hardware for a considerable sum.

Costs (EURO)	2004	2005	2006
project leader, e-conservator, ICT advisor	16.160	120.000	141.000
Staff GAR	26.000	49.563	49.563
courses	2.000	5.000	5.000
software adaptation	22.000	50.000	50.000
hardware (for 2005)	100.000	0	0
maintenance hardware	0	15.000	15.000
external expertise	18.500	19.000	19.000
PR, information	0	3.000	3.000
total	184.660	261.563	282.563

The above overview shows the time of Rotterdam Municipal Archives, not of the Archive School. It makes clear that costs have gone up in two years. Human costs are the biggest factor. Staff of Rotterdam Municipal Archives has given considerable time to the project in 2004 and 2005; consequences are that some core tasks of departments of the organisation will not be executed because of the project.

Results 2004 and 2005

In the time span of almost two years we have realised building stones for our digital repository:

1. Strategy: Rotterdam Municipal Archives is developing a strategy for long-term preservation;
2. Organisation: the implementation of the digital repository goes along with the development;
3. Staff: the learning process is evident; staff have developed competencies and knowledge and know how to describe gaps in knowledge;
4. Technique: we have the hardware; the functional specifications for the software are ready in 2005. Probably we will use existing applications of Rotterdam Municipal Archives as well for descriptive metadata.
5. Methods and documentation: In the modelling process according to Yourdon, we have developed and validated in 2004 and 2005 fundamental reports for our project:
 - a. The Organisation model: our starting point is custodial, no extra department Digital Conservation;
 - b. Environment model, Process model and Metadatamodel (all three models part of the Logical design = describes the functions of the system without the technology for these functions);
6. The results of the project are shared with colleagues in the Netherlands as it has no use to develop a digital repository alone. There is even practical participation in the project by some colleagues of other archival institutions.

Conclusion

The project shows that a digital repository project can start without huge costs; initially for starting one simple server or computer will do.

Open source products can help in the learning and developing as well, as they can be freely obtained. There are costs though of hiring open source expertise, unless one has specific IT knowledge about open source in the own organisation.

Obvious is the need to develop interfaces to facilitate transfer from the document management systems and other applications of the Rotterdam municipal organisation. That will be developed further in 2006.

In 2005 we started with letting colleagues from other institutions participate in the project: it is a win – win deal; they are informed about our results and progress and learn and we will have feed back and support.

The project drives on the involvement and commitment of all members. It will be successful if all relevant staff members have knowledge and skills necessary for the archiving of digital objects and feel the same need to preserve the digital as the non digital objects.