

Building a digital repository: a practical implementation

Filip Boudrez

Consultant digital archives

City Archives of Antwerp

Oudeleeuwenrui 29

2000 Antwerp

Belgium

filip.boudrez@stad.antwerpen.be

Abstract

Implementation of a digital repository involves much more than just the installation of the required hardware and software. It is clear from the implementation path undertaken by the Antwerp City Archives over the past year and a half that it is also an organisational issue and requires fundamental consideration of core records management and recordkeeping issues.

Introduction

Antwerp City Archives began the implementation of their digital archive repository in Autumn of 2007. Such a project however, involves much more than simply installing the required hardware and software. It is not only a matter of many organisational issues; the construction of a digital repository also raises fundamental questions about records management and recordkeeping.

1. The digital repository

Digital archiving has, for a long time, been a policy priority for the Antwerp City Archives. A proactive policy has already been in place for many years, focussing on city agencies and services that create records. During this time, the transfer of digital records to the archival service began and also many archival documents were digitised internally. Thus the volume of born-digital and digitised archives managed by the City Archives has gradually been on the increase.

The Antwerp City Archives anticipates that the digital repository will address procedures and infrastructure for the ingest, the management, and the dissemination of the digital archives and collections with medium to long term retention periods. The digital resources archived in the digital repository must be authentic and durable, regardless of whether they are born digital or digitised.

Construction of the Antwerp City Archives digital repository has, to a large extent, been based on the

research and recommendations from eDAVID.¹ Development and programming has, for the majority, been carried out in-house. Only the dissemination has been contracted out to an external developer, because this process is carried out separately through the website of the Antwerp City Archives.

2. An integrated recordkeeping system

An important and central aspect of the implementation is the positioning of the digital repository within the city archives. It is widely accepted that there are two main options for implementing this. One option is for the digital repository to be constructed alongside the recordkeeping system for analogue records and archives. This option results in only minimal integration between paper and digital records. The only integration envisaged in such instances is the use of metadata, to make the collections easier to search by researchers and civil servants working in the city administration. An alternative option is complete integration of paper and digital recordkeeping. In this case, the same procedures are used for treating and managing both paper and digital archives.

The City Archives has opted to implement option two for their digital repository and is therefore seeking maximum integration between paper and digital archives. There are many reasons for this. The first is so that the digital repository of the Antwerp City Archives may not be an isolated system. The second is that the City Archives wishes to implement a single system that will cater for all archives and collections, regardless of the medium or form of the records. The advantage of this is that the archival processes for intake, management and giving access are consistent for all types of records and only need to be automated once. The same software system can be used for managing all archives so that, in as far as is possible, information is shared and only needs

¹ See <http://www.edavid.be/eng/index.php>.

to be registered once. The principle of authority records is thus applied in a de facto manner and the need for developing similar functionality (for example, the ability to construct detailed retrieval aids) across different systems is avoided. Thirdly, integrated management of the archives is completely in line with the records creation practices that exist in a hybrid records creating environment. Finally, integration also contributes to embedding the digital repository within the institution.

The Antwerp City Archives is therefore constructing a single integrated recordkeeping system in which the management of paper and electronic records is fully part of. The integration takes place across numerous levels, including the software for processing and managing the records, the information architecture and the metadata model, and finally also the procedures.

3. Recordkeeping software: MACZ

Significant opportunities existed at the start of the implementation project that impacted on the choice of software for archives and records management. With the move to the Sint Felixpakhuis and the denBell project it was recognised that records management within the Antwerp City Archives was in need of revision. The new location and the application of box placement according to their size meant that a thorough and automated repository system was needed. Prior to the move, work also commenced on structuring and automating physical management practices in the archives.²

The denBell project started very soon after the move. This will soon result in administrative departments transferring large volumes of records to the archival service at pace. This must be done in an efficient way so they are quickly processed and accessible after their transfer. To facilitate this and ensure it is most effectively achieved, the way in which records are inventoried has been completely revised and the three international standards for archival description, ISAD(G), ISAAR(cpf) and ISDF, have been implemented. The way in which transfer lists were composed and processed has also been assessed and entirely adapted.

The new recordkeeping system has been named MACZ. MACZ will be completely managed and developed internally. As a result of opting for an integrated system, it was clear from the outset that the digital repository would be involved in the development of MACZ. The

archives will not therefore be managed by multiple and different information systems.

4. Information architecture and the metadata model

In revising the information architecture and data model for recordkeeping, paper-based records management was not necessarily assumed to deliver the leading principles. This wouldn't be strange as paper records management was a very familiar process. Instead, everything was thoroughly evaluated and optimum functional records management approaches and solutions were sought. Eventually it became apparent that the basic principles of digital archiving would provide the starting point for the new integrated information architecture and the new metadata model.

Two central principles for the new recordkeeping system in the Antwerp City Archives are, on the one hand, the separation of physical records management aspects from intellectual ones and, on the other hand, approaching archival records as abstract entities with one or more representations (paper, digital, microfilm/fiche etc). Separating physical and intellectual records management aspects has however a few consequences. The two aspects are managed and described separately but must remain clearly identifiable. Moreover, they must continue to be clearly linked to one another. In practice, this is achieved by linking the inventory numbers on the ISAD(G) records with descriptions of the records.

Another consequence was that ISAD(G) could not be implemented on its own and that the ISAD(G) implementation at the Antwerp City Archives deviated from the standard on numerous points. Diverse ISAD(G) description fields are related, for example, to representations of the archival documents (e.g. the element 'Extent and medium of the unit of description') and would be better described on the level of the inventory number. Such a ISAD(G) fields were not consequently implemented as proscribed by the standard but were given new functions.

This approach of identifying archival records using inventory numbers was already in place. This manner of working will be extended from paper record keeping to the digital archives. An inventory number is assigned to archival files and archival items as logical entities. For digital archiving a small amendment will be made: a single inventory number can refer to an entire series of digital archives to facilitate the complete retrieval and consultation of the full series. Request and retrieval of a digital series therefore has no logistical restriction in the way that paper archives do. In fact, this means that for the identification of a digital archive series an inventory number will be used in the meaning of ISAD(G)'s 'reference code' element

² See the annual report of the Antwerp City Archives (2006 – 07) for more information. Annual reports are available at www.felixarchief.be. More information about the denBell project is also available from this site, though in Dutch only.

Different representations of the same archival record are not allocated individual inventory numbers. On the level of the inventory number itself is registered which representations of the same documents is in the holding of the City Archives; for example, archival records in paper form have the inventory number linked to the physical storage box number (one or more); digital records will have a reference to their location within in the digital repository; and records on microfilm will be denoted by the microfilm number (one or more).

The further allocation of inventory numbers and their extension to digital records had, for the city archives, the significant advantage that the archives staff already was already familiar with this method and did not require any additional training. The manner in which digital records would be prepared for ingest into the repository was communicated over internal email. The new recordkeeping software supports the above in different ways. Furthermore, a built-in functionality enables members of staff to examine for themselves the suitability of the proposed digital inventory numbers.

5. New record keeping procedures and development

An important consequence of pursuing a completely integrated records and archives management system is that the procedures for paper and digital records management must be fully aligned. The practical actions for processing paper and digital records are naturally different, but will be embedded in the same basic procedures and follow a common workflow.

The original intention here was to take existing procedures for paper records keeping as the starting point. During this exercise it became apparent that many procedures for paper records keeping were not consolidated and that analogue archives were sometimes processed by archives staff in very different ways. The implementation of a structured and extensively automated recordkeeping system meant that it was consequently necessary to first ensure that consensus was reached between staff and that procedures were agreed across the board. This took a lot of meetings and time, and for this reason, the implementation of the digitale repository slowed down.

In parallel with planning the new recordkeeping and procedures, it was also necessary to programme and test the required software modules for the digital repository. The components of the different OAIS-processes – namely ingest, management, and dissemination - have been clearly defined and implemented. The functional model defined in the OAIS standard results in a basic workflow but this does not have to be strictly followed whilst developing and implementing the different

software modules. Instead, the essential components of each step of the process are developed simultaneously. For example, as soon as the first digital inventory numbers are ingested in the digital repository, the essential management quality controls must already be operational and the digital records must be retrievable and accessible. The ingest process does not therefore have to be fully implemented before development and implementation of the management modules is started.

This approach means that after the groundwork has been done on essential modules, work can begin on further refinement and optimisation. As a result, archives staff can already begin learning and gaining experience with the system. One advantage of this internal development is the ease and speed with which feedback can be given. The same approach will also be (more or less) followed to develop the dissemination functionalities retrieval and giving access. Until now, this is the only part of the digital repository that has not been developed by the City Archives itself but has been contracted out to an external partner as front-end access to digital or digitised archives is integrated into the website of the City Archives. Development of the front-end portal is happening in two phases. During the first phase, the basic modules and interfaces will be developed. These will only be available to archives staff, who can retrieve and consult the digital archives. During the second phase, the remaining functionality will be implemented (registration of consultations, security etc) and will be further optimised according to feedback. The request process will be refined according to feedback from users and will be made more intuitive for the end user.

Gaining experience is valuable for both archival staff and external developers. A digital repository differs in numerous respects from other types of computing systems. Electronic records can have very complex structures (for example, a website), or can be extremely bulky (for example a digitised register). Such factors must be taken into account when the technical issues relating to query and consultation functionality are being worked out.

Alongside development of the first modules, the storage infrastructure was installed and configured. Digital objects will be securely and safely stored in a SAN (Storage Area Network). The available storage capacity of the digital repository will be systematically expanded on a step-by step basis. This step-by-step approach will safely support tests of future development activities to increase storage capacity. Additional and particular points of interest included definition of the security protocols, organisation of the back-up process, and the integration of the storage infrastructure. The storage environment must be secure not only against regular users but also applications such as those used within the front-end access portal. The digital repository differs from regular digital information systems, and this even extends to the backup systems. Due to the high volume

of backup data, a standard backup regime is insufficient. The creation of a full backup is a time consuming activity and cannot be carried out weekly or even twice a month. Nevertheless, the creation of good backups remains an essential part of a disaster recovery strategy.

Concern over this last point and a supporting risk analysis resulted in significant attention being paid to the arrangement of the storage infrastructure, which was not selected for technical or efficiency reasons but for providing support to meet minimum archival requirements/logic. Should data be lost at the database level or subsequent references become incorrect, the essential linkage and management data can quickly be generated anew or corrected.

6. Transferring the records into the digital repository

As soon as the essential modules for ingest and management were operational in the storage infrastructure, transfer of items into the digital repository could commence. By 2007, the Antwerp City Archives already had collected 1TB of digital and digitised records. These digital records had until then mainly been kept on CD, with some exceptions (f.i. preserved websites) stored on file servers. Transferring these legacy files was thus one of the first tasks.

As a result of the sheer amount of work and the additional activities this task entailed, it was something of a sub-project in its own right. Just copying the contents of the CDs had already taken a long time. Moreover, the contents of the CDs could not simply be transferred direct into the digital repository. In most cases, one or more archival processes had to be undertaken upon the digital records: assigning inventory numbers, abolishing inventory numbers not accepted by the new system and therefore neither by the digital repository, adapting folder names, checking integrity, registering metadata that had been distributed up until then, and so forth. For the most part, these archival processes generally required additional and meaningful checks or analysis so dealing with them took a considerable number of months.

As soon as the first digital inventory numbers were ingested in the digital repository, essential management tasks had to be carried out. This assumed in the first instance that responsibility for management of the digital repository had already been clearly determined. It was expected that the digital repository manager would be thoroughly acquainted with the metadata and database models. This was already the case for the Antwerp City Archives, for not all of the required user interfaces and associated modules for managing the system were ready and a certain amount of work still took place behind the scenes.

A particular activity in this part of the implementation process involved the AVA image database. AVA contains descriptions of digital photographs and makes them available online. Master copies are now stored in the digital repository, whilst low resolution copies are uploaded to AVA for access. The digitised photos are therefore managed in both systems, which means AVA must be integrated with MACZ and the digital repository. However, as AVA was developed externally it has not been possible to integrate all of the functionality in with the recordkeeping system. Eventually it became clear that the most important integration work would be realised within the MACZ environment. The AVA database itself needed only one minimum adaptation. Once this adaptation was complete, work could commence to transfer the AVA master files into the digital repository.

Implementation of the digital repository has led to questions being asked about the future of the AVA database and its content. The digital repository enables direct access to digital photos via the City Archives' website without needing to access AVA itself. Numerous photos and pieces of iconographical work that had been previously stored in AVA are no longer stored there as these contents have been transferred to the digital repository. The AVA records will be deleted once the descriptions have also been extracted.

7. Conclusion

From the implementation path followed by the Antwerp City Archives, it is clear that the construction of a digital repository is initially more concerned with structured and efficient record keeping than with digital archiving in particular. Consequently, the implementation of a digital repository should ideally begin with a records management and recordkeeping 'vision' in which the best parts of paper and digital records management are brought together. Once the general procedures have been established for ingest, management, and making archives available, implementation of the digital repository should proceed relatively smoothly.

Implementation at the Antwerp City Archives has taken place in a step-by-step manner. Because of this it has been possible to learn about and systematically align records management practices, and thus also improve the digital repository. Implementation is therefore also an iterative process, in which consolidation and improvement continuously alternate with each other. The digital repository remains thus a work-in-progress.