

# Filing and archiving e-mail

Filip Boudrez  
Expertisecentrum DAVID vzw  
Antwerp, 2006

## 0. TABLE OF CONTENTS

---

1. Introduction.....	1
2. Quality requirements for an e-mail archiving procedure.....	3
2.1 Judicial framework.....	4
2.2 Archival and organisational requirements.....	5
2.3 Implementation criteria.....	6
2.4 The DAVID model solution.....	7
2.5 Market investigation and evaluation.....	13
3. Filing e-mails and electronic documents.....	14
3.1 Building a classification system and creating electronic files.....	14
3.2 Registering metadata.....	16
3.3 Filing e-mails and attachments.....	19
3.4 Customisations.....	21
3.5 Implementation.....	32
4. Archiving electronic records.....	35
4.1 Selection of the files with archival value.....	35
4.2 Archiving metadata.....	35
4.3 Migration to preservation formats.....	37
4.4 Encapsulation in AIP's.....	39
4.5 Retrieval and dissemination.....	40
5. Conclusion.....	41
6. Appendices.....	42
6.1 Tools.....	42
6.2 Alternative implementations.....	42
6.3 Roles and responsibilities.....	44
7. Abbreviations.....	46
8. Literature.....	46

## 1. INTRODUCTION

---

*Preserving and  
archiving e-mail*

E-mail systems have become so firmly established for the communication of information that they have gained the status of business-critical applications. E-mail not only brings a faster and more efficient exchange of information, but also new challenges in the areas of records management and record-keeping. Oversized mailboxes, unreadable e-mails, and losing time while searching e-mails and related documents are problems that everyone recognises. E-mails are a good example of a new technology which results in records creation, records management challenges and record-keeping issues. E-mails, and electronic documents exchanged by e-mail, can have record-status or regulations regarding freedom to information might be applicable, and are then eligible for medium-long or long-term archiving. Therefore, administrations and archivists must certainly

deal with the management and preservation of e-mail. The DAVID project<sup>1</sup> examined the judicial and archival requirements for e-mail preservation and pointed out some possible archiving strategies (Report 5<sup>2</sup>). On this basis, a model solution was developed. In addition to the theoretical concept, this report also contained an initial incentive for the practical implementation of a records management and record-keeping procedure for e-mails and related electronic documents.

*A practical solution* This present report builds further on the DAVID study of e-mail archiving. It first indicates how e-mails can best be managed and archived, and secondly how the Antwerp city archives developed a custom-made records management and record-keeping procedure for e-mails and attachments for the city administration of Antwerp and how it is putting all this into practice. The city administration has more than 6500 users of e-mail. A practical, scalable and user-friendly translation of the DAVID model solution was sought for the agencies of the city administration. These implementation criteria are important in order to have maximum compliance with the outlined procedure. This led to the development of an archiving procedure that runs from the creation or receipt of e-mails to the retrieval of archived e-mails.

Implementation started in 2002. The procedure and some prototype instruments were tested by pilot projects in the municipal agency for human resources. The experience gained led to several adjustments in the area of user friendliness. For the practical implementation, the necessary software tools were programmed. All of these instruments have been developed by the DAVID project and the Antwerp city archives.

*Electronic records management and record-keeping* The second central theme of this report is the opportunity that e-mail archiving offers an organisation for putting electronic records management and record-keeping on the agenda and into practice. The records manager or archivist can use e-mail archiving as trigger to do something about the management and archiving of electronic documents in general. In addition to e-mails and their attachments, organisations also have many other electronic office documents that are kept at various locations. An archiving strategy is needed for these electronic documents as well. An efficient strategy for filing and archiving e-mail should be correlated with the general electronic-records management and record-keeping of the organisation. If one does not exist, e-mail archiving can be a good occasion for developing one. The Antwerp city archive incorporated their archiving strategy for e-mails and attachments into the overall archiving procedure for electronic office documents. This report goes farther in this regard than the DAVID report and also describes the following steps in the archiving process: migration to suitable preservation formats, transfer to the archives, ingest in the repository, archival description, retrieval and dissemination.

*Incorporation into the existing IT environment* The archiving procedure is elaborated within the existing IT configuration. This is a conscious choice. In this way the administrative employees and the archivist continue working in a software environment with which they are already familiar. This option also shows that without large additional investments, a number of important steps can be taken with regard to electronic records management. The

---

<sup>1</sup> DAVID means 'Digitale Archivering in Vlaamse Instellingen en Diensten' [Digital Archiving in Flemish Institutions and Departments] and was a 4 year funded research program with the Antwerp city archives and the Interdisciplinary Centre for Law and IT (University of Louvain) as project partners.

<sup>2</sup> F. BOUDREZ, H. DEKEYSER and S. VAN DEN EYNDE, *Archiving e-mail*, Antwerp-Leuven, 2003 (Version 2.0).  
([http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Rapport5\\_V\\_2.pdf](http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Rapport5_V_2.pdf))

city administration of Antwerp uses Microsoft Exchange and Outlook as e-mail system. Institutions or organisations working with different e-mail-server or e-mail-client software (such as Domino - Lotus Notes, Eudora, GroupWise, Thunderbird) can draw inspiration from this report and work out an analogous solution. The commonly used e-mail systems all have similar basic functionalities. For the management of electronic documents in general, use is made of Windows Explorer. Several city agencies are working with Documentum and Docushare as records management application, but they are rather the exception than the rule. Furthermore, the same basic principles apply for the organisation of records in digital series and files, whether they are stored in an ordinary file-based file system of operating systems or a more advanced records management application.

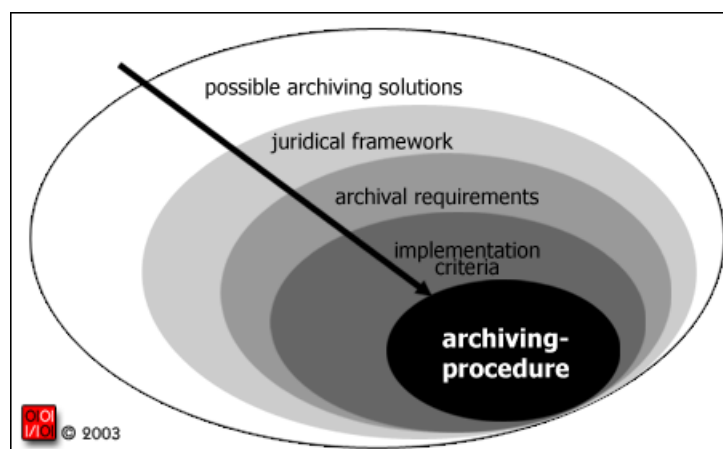
*Structure of the report*

This report consists of three main parts. First, the general quality requirements for an archiving strategy for e-mails and attachments are described. This includes the legal framework, the archival requirements and the implementation criteria. Within these guidelines, an archiving strategy is developed. How these quality requirements were translated into a records management and record-keeping procedure for the city of Antwerp is discussed next. Electronic records management has the main focus in the second part of this report. Emphasis is placed on the filing of electronic documents in general and e-mails in particular. Attention is given to practical implementation and the instruments used. In this section, several technical aspects of e-mail archiving are discussed in greater detail. And finally, in the last part, the long-term preservation of electronic documents is discussed. Electronic records with archival value are prepared for transfer and are ingested in the digital repository.

## 2. QUALITY REQUIREMENTS FOR AN E-MAIL ARCHIVING PROCEDURE

*From possible solutions to an archiving procedure*

In the fifth DAVID report, the general judicial and archival framework for archiving e-mail was outlined. This study defined the borders within which an archiving procedure for e-mails and their attachments can be developed.



**Illustration 1:** After testing the possible archiving solutions against the judicial framework, the archival requirements and the implementation criteria, an archiving procedure is defined.

## 2.1 Judicial framework

*Preservation and archiving is a legal obligation* The legislator obligates public institutions to archive e-mails and also defines the limits within this might occur<sup>3</sup>.

The government has an obligation to retain and archive e-mails with record status and/or e-mails to which the freedom of information act is applicable, in a good, orderly and accessible state. This obligation emanates from archival legislation and the freedom of information act. Both laws provide the public sector with a basis for e-mail archiving as a legitimate objective, but one must be careful that private e-mail is kept out of the archive and that the rights of the e-mail users are not violated.

*Legal barriers* The limits within which an organisation can act, are determined in particular by the protection of privacy, by freedom of communication and by telecommunication secrecy, all of which are based on art. 8 of the European Convention on Human Rights (ECHR)<sup>4</sup>. The principles established in art. 8 of the ECHR are further defined in Belgian legislation by the law on the protection of privacy and the provisions regarding telecommunication secrecy. In Belgium, the concept of privacy is interpreted very broadly as professional communication is also protected by this legislation. According to art. 8 of the ECHR, an employee has the right to make use of the communication resources of the employer, also for private purposes. This right is not unlimited, but the employer may not absolutely forbid the use of e-mail for private purposes. Telecommunication secrecy forbids all interference in the correspondence or exchange of information between other persons. Gaining knowledge of the existence and of the content of telecommunication is in principle punishable by law. Even making a copy without opening the message, is covered by this. The exact scope of this prohibition has been contested for years, but recent jurisprudence limits the protection strictly to the transfer phase. According to this interpretation, the impact of telecommunication secrecy on the archiving of e-mail is rather small, but not non-existent. Because the legislator is aware that this rule can come in conflict with other interests, several exceptions are provided. First, archiving is not punishable if the archivist has the permission of all participants in the communication. This basis for an exception cannot be used for the archiving of all e-mails with record status, however, because then the approval of the sender and all addressees would be required each time. Second, archiving that is required or allowed by law is not punishable. The legal obligation to archive records and administrative documents falls within this exception.

*The law on privacy* The law on privacy also applies to e-mail. This law regulates the processing of personal data. Almost every e-mail contains personal data and must be treated in accordance with the principles of this law. Completely automated, systematic archiving of all e-mails by the employer is considered to be an encroachment on the privacy of the employees. The law on privacy allows this encroachment only if three principles are respected: transparency, finality and proportionality. Transparency means that all involved parties must be informed about the archiving policy. E-mail may only be archived in the framework of a legitimate objective, for example, the legal archiving obligation or the obligation to make records accessible and public.

---

<sup>3</sup> With thanks to Hannelore Dekeyser for bringing this chapter up to date. For more information about the legal framework, see: F. BOUDREZ, H. DEKEYSER and S. VAN DEN EYNDE, *Archiving e-mail*, Antwerp-Leuven, 2003 (Version 2.0).

<sup>4</sup> The legal basis for this is: the Belgian Constitution, art. 22 and 19; Law of 21 March 1991 concerning the reformation of certain economic enterprises (Belgacom or Telecom law), art. 109terD and 109terE; Law of 8 December 1992 on the protection of privacy.

The processing of personal data must be in proportion to this legitimate goal, which is why only professional e-mail may be included in the archive.

*Only professional messages may be archived*

The organisation may archive e-mail to the extent that it relates to e-mails with record status or to which the freedom of information act is applicable. Private e-mail may not be archived by the employer. To make a distinction between private e-mail and professional e-mail, the co-operation of the end user is the only workable solution. Automatic and direct archiving by the e-mail server without the intervention of the end user is not allowed legally. The organisation must formulate clear rules for the processing of e-mail by the end user, namely, in order to separate work-related and personal e-mail. This can be put into practice by having the employee add the e-mail to a file or forward it to a records manager who then takes care of files management. In this way private mails are separated from e-mails that relate to business or subjects of the organisation, and these e-mails are no longer opened or registered during their transfer.

## **2.2 Archival and organisational requirements**

An archiving strategy for e-mails and attachments must be drawn up within this legal framework. The archivist must also take archival needs into consideration and several criteria for successful implementation and application.

*Archival context*

First, just like all other electronic records, e-mails and attachments must be archived within their archival context. E-mails, along with their attachments, must be interpretable in future. They must therefore remain related to their creator and situated in the work process in which they were created or received. In future, the series, the file or subject to which an e-mail relates must be clear. The mutual relationships among records that belong together must also be preserved. This applies, not only to the association between an e-mail and any attachments, but also to the relationship with other paper and electronic documents in the organisation that relate to the same file or the same subject. This has two direct consequences for the archiving strategy. First, only e-mails and attachments with record status must be archived. Second, documents with this status must be situated within their context and this relationship must be preserved. For that reason it is best to link the records to their context in an explicit way. For the archiving strategy this means that intervention is required by people who are very familiar with the function and the meaning of the e-mails and attachments. The person in the organisation who is best placed for this is the sender or the recipient of the e-mail message.

*Essential components of e-mail*

The authenticity of archived e-mails also requires that all essential components of an e-mail be archived. In addition to the archival context, the content and the structure of the e-mail message are also essential. The content of an e-mail includes not only the subject description and the message field, but also any attachments<sup>5</sup>. The internal structure reflects the relationship among the components of an e-mail: header data, message field, attachments. An e-mail is only complete when all of these components and their mutual relationship are preserved. In general, behaviour and layout are not included among the essential

---

<sup>5</sup> Moreq: 6.1.2

components. E-mails are, after all, static and do not have a unique layout. The layout of an e-mail message is dependent on the client software used<sup>6</sup>.

*Essential transmission data* In addition to the context and the components mentioned, several items of the e-mail transmission data must also be archived. These transmission data can be viewed as metadata. There is a general consensus about which intrinsic elements are needed for the identification of an electronic record as an e-mail<sup>7</sup>: the unique ID, the name and the e-mail address of the sender, the date and the time of sending, the name and the e-mail address of each addressee (To, CC, BCC), the date and the time of receipt, the subject, and the number of attachments. These data characterise an e-mail and distinguish it from other documents. Most of these transmission data are found in the e-mail header.

*Long-term readability* Third, e-mails and their attachments must be archived in a permanent way. To anticipate the digital readability problem, an attempt is made to be independent from any specific hardware and software as much as possible. The electronic records are therefore archived in a platform-independent manner. Not only the e-mails and the attachments, but also their context and archival bond have to be permanently preserved.

*Embedding within the organisational context* The archiving strategy must, in the forth place, be embedded into the organisational context of the institution. Which e-mails have record status for the organisation? In which work processes are e-mails sent and received? How is the archiving of paper and/or electronic documents organised in general? What is the technological infrastructure of the organisation? How are the authorisations and responsibilities distributed regarding to records and IT management?

### 2.3 Implementation criteria

*User-friendliness and easy deployment* Finally, practical and scalable solutions are sought. It is preferable to deploy archiving solutions within the existing IT infrastructure. Then large investments are avoided (additional software licenses, training courses, etc.) and the user can continue working with computer programmes they are already familiar with. Together with a practical and simple procedure, this should contribute to a very good applied archiving procedure. Automation should be used whenever it's

<sup>6</sup> The formatting of e-mail messages must be viewed as an extension of the e-mail standard. Many e-mail clients, for example, do not support HTML and RTF formatting of the message field. Thunderbird, for example, automatically converts RTF formatting to HTML. Certain versions of Netscape Messenger convert HTML layout to plain text. For this reason, some programmes, such as Thunderbird, allow you to set which addressees do or do not have HTML capabilities.

<sup>7</sup> RFC 822, *Standard for the format of arpa Internet text messages*, 1982; RFC 2822, *Internet Message Format*, 2001 (<http://www.ietf.org/rfc/rfc2822.txt>); DoD, *Design criteria standard for electronic records management software applications. DOD 5015-2*, 2002, p. 32-33 (C2.2.4.2); TESTBED DIGITALE BEWARING, *Van digitale vluchtigheid naar digitaal houvast*, The Hague, 2003. p. 26ff, INTERPARES 1, *Template for analysis*, 2000. Moreq and ReMaNo do not consider which transmission data are essential for an e-mail and therefore must be recorded in an RMA. Moreq and ReMaNo only state that it is preferable to retain the name of a correspondent written in full rather than an e-mail address (Moreq: 6.4.3; Remano: 162). In the Dutch translation this is translated as the 'interpretable version of an e-mail address' whereas the name of an account identity is actually intended. In Moreq and ReMaNo an e-mail address in SMTP style is assumed, although e-mail addresses can also have an X.400 style. Moreq does state that the transmission metadata of an e-mail should be protected against modifications (Moreq: 12.1.23).

possible. This limits human intervention, avoids human mistakes, contributes to user friendliness and assures a good application of the archiving procedure. In addition to the judicial and archival requirements, this pragmatic approach will influence the selection of a certain archiving strategy. Scalability is a factor that must be given special consideration in large organisations.

## 2.4 The DAVID model solution

*DAVID research* Archiving e-mails was a specific research area within the DAVID project. Within the designated judicial and archival framework, an archiving solution for e-mails was sought. Organisations which want to develop a custom-made archiving policy for e-mails (and electronic documents) can use this model solution as a basis. The general DAVID approach for e-mail archiving can be implemented in various ways and in different technological environments. The DAVID strategy is designed to preserve usable e-mails, attachments and other electronic documents. This means that the documents are retrievable, readable and understandable<sup>8</sup>.

The following steps are part of the DAVID model solution:

- 1) registration of the transmission and contextual metadata
- 2) electronic filing: exporting e-mails and attachments and keeping them together with related documents
- 3) migration of e-mails and attachments to preservation formats

### 2.4.1 REGISTRATION OF THE TRANSMISSION AND CONTEXTUAL DATA

*Registering transmission data* The essential transmission data of e-mails are: a unique ID, the name and the e-mail address of the sender (and his authorised delegate), the date and the time of sending, the name and the e-mail address of the recipient(s), the date and the time of receipt and the number of attachments. These data are present in the e-mail system for each e-mail but they are not always shown together and they sometimes change (for example, through dynamic retrieval of e-mail addresses from the address book). For the sake of the completeness and the authenticity of the e-mail as a record it is important that all of these data are registered in a structured and static manner and are inextricably archived together with the message. The best method for this is the embedding of these data so they become an internal part of the e-mail. This is also an important point for consideration when e-mails are preserved on paper<sup>9</sup>.

*Indication of the context* To ensure the future interpretation and understanding of an archived e-mail, one must know the context within which the e-mail was used. The relationship between the e-mail, on the one hand, and the creator and the work process on the other hand, must be indicated in one way or another so the meaning and function of the record can be discovered. This can be accomplished by using the filing code or by adding another registration reference to the e-mail. These descriptive metadata

<sup>8</sup> ISO-15489 defines a usable record as “one that can be located, retrieved, presented and interpreted” (ISO-15489: 7.2.5).

<sup>9</sup> Preserving e-mails on paper (= the hard copy option) is not dealt with extensively in this technical report. For this see: F. BOUDREZ, H. DEKEYSER and S. VAN DEN EYNDE, *Archiving e-mail*. Antwerp-Leuven, 2003 (Version 2.0).

should indicate the context and also the finding place of the document. Since such a reference establishes the archival bond, this is an important identifying component of the e-mail as a record. The status of 'record' depends, among other things, on that reference to the context.

*Who registers the metadata?* As the essential transmission data are present in the e-mail system, they can be retrieved and recorded completely automatically for each e-mail with archival value without any action being required on the part of the end user. The assignment of a filing code or registration reference, however, cannot be done completely automatically. The sender or the addressee is in the best position to know the context in which a message was received or sent, and is therefore the best person to contextualise a message by assigning it to a certain dossier or folder. It is important for this operation to be as user friendly – and efficient as possible. Automation can be a big help in this.

*When should registration occur?* Preferably, both the transmission and the contextual metadata should be recorded when the e-mails are still in the e-mail system. Ideally, the 'capture' moment should be as close to the time of sending or of receipt as possible.

## 2.4.2 ELECTRONIC FILING: EXPORTING E-MAILS AND ATTACHMENTS AND KEEPING THEM TOGETHER WITH RELATED DOCUMENTS

*Filing and classification* The e-mails and attachments are arranged and organised in folders. For this, a folder structure is constructed in which e-mails and attachments are stored and can be retrieved when needed. The folder structure of the electronic filing system makes the structure of the archive visible and integrates the documents with their work process. The e-mails and attachments are grouped within the folder structure per file or subject. Thus dossiers and folders are created and arranged according to a certain logic. Ideally, the construction and hierarchy of the folder structure should be based on the tasks and activities of the creator. Not only is this commonly considered to be the most stable classification criterion for records, but in a classification system based on tasks or operational processes, electronic documents will retain their full meaning and will be (re)useable. Information about the context of the filed and/or archived e-mail and attachments is then communicated by the folder structure and the place of the electronic records and files within that folder structure. The electronic documents are then directly linked to the operational process in which they had a function<sup>10</sup>.

*Why exporting e-mails and attachments?* Commonly-used e-mail systems provide the possibility of creating an on-line or an off-line folder structure, and of moving e-mails and attachments to those folders. A folder structure in the e-mail system is, however, only suitable as a temporary storage place, and certainly not as the final destination of e-mails and attachments with record status. Export of e-mails and attachments to a folder structure outside the e-mail system is required for several reasons.

---

<sup>10</sup> A classification schema based on the business processes and the tasks or operational processes is central in DIRKS and in ISO-15489 (the standard DIRKS inspired). The essential characteristics of a 'record' are determined on the basis of the operational processes. (DIRKS stands for 'Designing and Implementing Record Keeping Systems': <http://www.naa.gov.au/recordkeeping/dirks/dirksman/dirks.html>)

First, there is the digital durability problem. Most e-mail systems use their own file or database format for storing e-mails. On-line and off-line folders are usually compressed computer files or small proprietary database applications, which can cause readability problems as time and (versions of) applications goes by<sup>11</sup>. Therefore it is best not to use the 'archiving' functionalities that certain mail software packages provide. These functionalities are mainly designed to reduce the load on the e-mail server and to temporarily put e-mails and attachments aside in closed and compressed files.

Second, e-mails in the e-mail system are not always easy to access: mailboxes and off-line folders are protected by accounts and passwords, off-line folders are difficult to share with colleagues, etc.

Third, e-mail systems and their storage facilities are not suitable for the management of large quantities of e-mails and attachments. Large on-line folders impair the performance of the servers, while off-line folders, because of their large size, easily become corrupt and are therefore unreliable and unstable.

Forth, when e-mails are exported, the link with the mail server is broken. This has the advantage that certain items of information, such as e-mail addresses, are no longer automatically modified (for example, after updating the address book) and are therefore static.

The fifth reason for the export of e-mails and attachments is the integration with related electronic records that are not sent through the e-mail system. It is not easy to include such records in the folder structure of an e-mail system. Yet, they can relate to the same file or subject and they should therefore be preserved together with related e-mails and attachments. The reverse, however, is easier to accomplish: e-mails and attachments can be moved outside the e-mail system and preserved together with the other electronic documents of the organisation. By preserving all relevant documents together, an overview of a file or a subject can be reconstructed faster and more accurately afterwards. Thus, the folder structure designed for e-mail archiving also provides the possibility of preserving other electronic documents in a structured way and in their context. Material at the various storage locations for electronic documents within the organisation (e.g. e-mail system, file servers, local hard disks) can then be moved to one shared folder structure, which increases the opportunities of finding, sharing and reusing existing information. By integrating e-mails and attachments with the other electronic office documents, electronic files are created that are kept at a central place. Centralised administration offers advantages in the area of management (security, back-up, accessibility, etc.). This is an important step on the way to controlled and structured records management.

And finally, exporting e-mails and attachments also has the practical advantage that the filed e-mails and attachments remain available when the e-mail server is not accessible.

---

<sup>11</sup> The MS Exchange and Outlook environment is a good example of this. In MS Exchange and Outlook, the storage places of mails are in on/off-line folders and post boxes, in the Exchange Information Store databanks, and in Outlook \*.pst files. The databases of the Exchange Information Store are saved on one or more servers. Outlook \*.pst files are usually preserved on local hard disks or server disks. In the case of an open-source e-mail client such as Thunderbird, the format of the local folders is documented, but it is not a suitable archiving format.



*Business cases:  
the opposite  
approach* Some e-mail archiving solutions are based on the opposite method, however, with an electronic filing system being developed within the e-mail infrastructure. Especially in the private sector this approach is often applied. The user-friendly and more sophisticated search possibilities of an e-mail client programme such as MS Outlook or Lotus Notes are put forward as an argument for this. For the above-mentioned reasons, however, this approach is not recommended. E-mail systems are, after all, not records management applications. Furthermore, such an approach involves other electronic documents being imported into the e-mail system, even though they were not received or sent by e-mail.

*Which e-mails  
and attachments  
should be filed* Only e-mails and attachments with record status for the organisation belong in the electronic filing system. Personal e-mails, e-mails of a purely informative nature, etc. should not be preserved in the electronic filing system of the organisation.

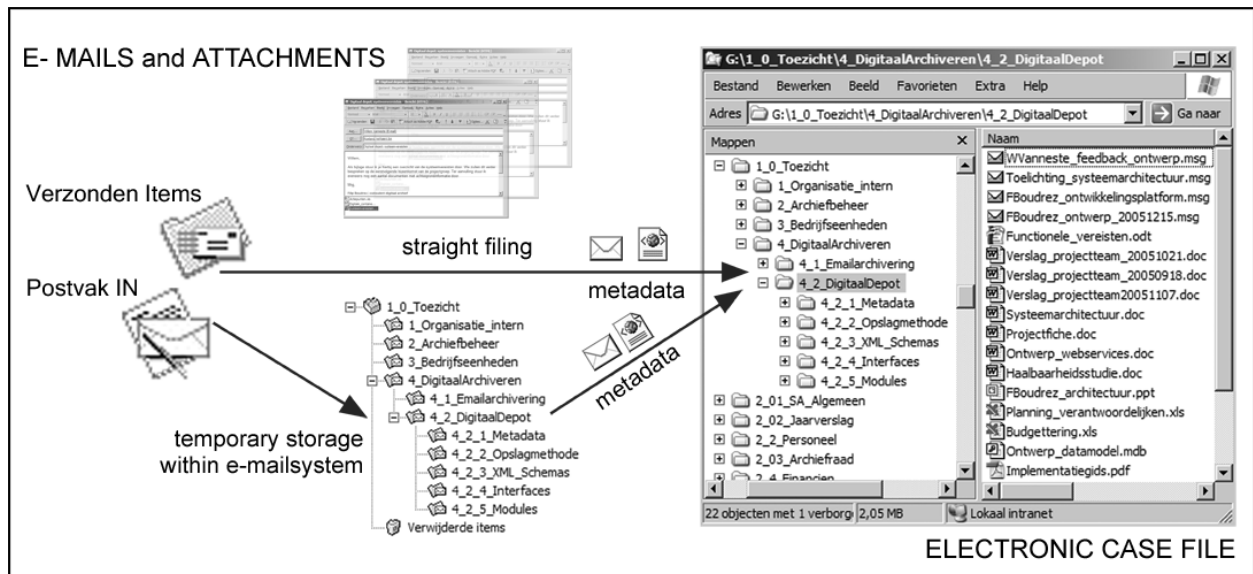
Selection is also urgently needed in a digital environment. Although commercial players on the archiving market have promoted the opposite for years, the most recent generation of archiving solutions starts with the need for selection. Archiving everything is not only very expensive but also increases the search time significantly, even if one has access to automated search technologies<sup>12</sup>.

*Who files e-mails  
and attachments* The sender or the addressee is the most obvious person to file e-mails. There are both judicial and archival reasons for this. Allowing the sender or the addressee himself to decide whether to file his e-mails is the safest way to avoid encroachment on the privacy of the employees. From an archival point of view, the end user is in the best position to judge whether the e-mail message and/or the attachments are records, and if so, to indicate the series or file to which they belong.

The intervention of the end user is an important success factor. This of course involves several risks, such as insufficient compliance with the archiving procedure, the development of a personal filing system outside that of the organisation, or the wrongful deletion of records. One must take this into consideration when developing a concrete deployment and implementation procedure. In the practical application it is also advisable to make clear agreements within the organisation as to who files an e-mail message that was sent to several addressees.

---

<sup>12</sup> In the commercial world this is called the 'big-dump' approach: 'archive everything and hope for the best'. Practical experience indicates however that this results in large volumes of poorly indexed e-mails and labour-intensive searches (D. REIER, *I Have to Show Them What?! E-Mail and the process of electronic discovery*, in: *Information storage and security journal*, June 2005).



**Illustration 2:** Creating electronic files by exporting e-mails and attachments, and grouping them with related documents. E-mails and attachments can be preserved temporarily within the e-mail system or can be moved immediately to the appropriate electronic folder in the classification system.

### 2.4.3 MIGRATION OF E-MAILS AND ATTACHMENTS TO PRESERVATION FORMATS

Archiving e-mails as XML documents

Before e-mails and attachments with archival value are ingested in the digital repository, it is best for them to be migrated to a suitable preservation format. Since e-mails are well-structured and are textual documents, XML is the obvious choice for the long-term preservation of e-mails.

XML<sup>13</sup>:

- is an open standard of the *World Wide Web Consortium*. The XML specification is stable, open and public. The specification can only be changed after going through a whole procedure and after consultation with various parties including the public.
- is free of patent and licensing rights
- is platform independent. An XML document is in essence nothing more than a flat text file (Unicode) that can be consulted with various software applications. For long-term archiving, textual encoding is also safer than binary encoding<sup>14</sup>.
- separates layout from content and structure. An XML file contains the content and the structure of a document. The layout of a document is defined with a stylesheet (CSS, XSL)<sup>15</sup>.

<sup>13</sup> For a more complete overview of the advantages of XML for archiving purposes, see: F. BOUDREZ, <XML/> and electronic record-keeping, Antwerp, 2002 ([http://www.expertisecentrumdavid.be/davidproject/teksten/XML\\_erecordkeeping.pdf](http://www.expertisecentrumdavid.be/davidproject/teksten/XML_erecordkeeping.pdf)).

<sup>14</sup> One error in a binary file can lead to the permanent loss of a complete record, whereas with textual encoding the rest of the record can still be reconstructed.

<sup>15</sup> The stylesheet can be stored within the XML-document (e.g. for dissemination) or in a separate file.

- is extremely suitable for transferring a document model through time in an explicit way because of the combination of nesting and semantic tags. Since XML is extensible, the user can employ his own document models.
- can preserve the structure of an e-mail in an explicit way within the document itself. This makes it possible to do structured searches on the header fields, for example. The structure is also documented externally in a DTD or an XML Schema.
- offers several validation possibilities so the quality of the XML documents can be checked automatically
- has great market penetration
- is an exchange format that is suitable for becoming the basic format for e-mail transmission<sup>16</sup>.

Since, at present, e-mails are still communicated as regular flat text files, a migration must be provided for the XML preservation of e-mails. This migration consists mainly of the addition of XML tags to the various data fields and the structuring of the intrinsic e-mail elements. Since commonly-used e-mail systems are not yet equipped with such a functionality, an ad hoc solution is needed for this. One can use a separate computer programme for the migration, or incorporate such a functionality into the e-mail programme (see further).

*PDF/A as an alternative*

An alternative for XML as the archiving format is the PDF/A format that has been established as an international standard (ISO 19005-1:2005). PDF/A is intended as a limited but stable subset of the PDF format of manufacturer Adobe. PDF/A provides several advantages compared to PDF. PDF/A is a standard for textual documents, of which the management is no longer in the hands of one company, but of a standardisation agency in which the government, the manufacturers and the academic world are represented. This guarantees greater stability and certainty. Adobe controls PDF completely on their own and are not at all obligated to publish the PDF specifications. PDF/A has been specifically constructed for archiving purposes. PDF/A documents must be self-reliant and must avoid external dependencies (such as the retrieval of external fonts, or encryption) and proprietary applications as much as possible<sup>17</sup>.

*Preservation formats for the attachments*

To determine which file formats is a suitable preservation format for e-mail attachments and other electronic documents, consideration is given to such things as the type of document, its characteristics and the way it's used within the creating agency. Each type can have specific archiving requirements both in the area of suitable preservation formats and of metadata. This is one of the reasons that e-mails and attachments are separated when they are moved outside the e-mail system. Digital Archiving: guideline & advice no. 4<sup>18</sup> provides an overview of suitable preservation formats for various types of electronic documents.

<sup>16</sup> See among others G. KLYNE, *An XML format for mail and other messages*, 2003. This is a proposal for e-mails to be encoded in XML in conformity with RFC822.

<sup>17</sup> For more information about the PDF and the PDF/A formats: F. BOUDREZ, *Standaarden voor digitale archiefdocumenten*, Antwerp, 2005. ([http://www.expertisecentrumdavid.be/docs/eDAVID\\_standaarden.pdf](http://www.expertisecentrumdavid.be/docs/eDAVID_standaarden.pdf))

<sup>18</sup> <http://www.expertisecentrumdavid.be/davidproject/teksten/guideline4.pdf>

## 2.5 Market investigation and evaluation

*Business cases* Existing solutions were evaluated before our own archiving procedure was developed and the associated tools were programmed. Several archiving solutions from the private sector were tested on the basis of the judicial and archival requirements, but didn't comply. The lack of contextual information and of a vision for long-term archiving are the main reasons for this (see also 2.2: Business cases: the opposite approach).

*Commercial applications* In addition to business cases, the commercial market was also investigated. The main players on the e-mail archiving market were invited to present their products. Digipolis, the information-technology partner of the city of Antwerp, and the city archives tested the proposed commercial archiving solutions against the designated technical, judicial and archival quality requirements. The products KVS and Enterprise Vault, Email Archive Manager and Exchange Archive Solution were evaluated. These products all provide the same basic functionality: during archiving, the e-mails and attachments are moved from the e-mailserver to a separate server. In the mailboxes, the archived e-mails are replaced by shortcuts so the load on the mail server is reduced. From a database, the archived e-mails and attachments can still be retrieved.

Not a single one of the proposed commercial products met the preconceived requirements. General shortcomings of these commercial packages are<sup>19</sup>:

- direct archiving on the e-mail server during the transmission phase and without the intervention of the end user, which is difficult to accomplish within the Belgian legal context.
- limited filing functionalities: only electronic documents sent by the e-mail system can be filed in the electronic classification system. Electronic documents that are not sent by e-mail, cannot be added to the filing system.
- loss of archival context and related retrieval / browse functionalities. The folder structure cannot always be taken over. The retrieval added-value of certain storage systems in the form of full-text searches does not compensate for the loss of archival context and browse possibilities based on the folder structure and on contextual header data.
- no central or co-ordinated records management: the logical organisation of the e-mail archive is left to the user who manages his mailbox himself with shortcuts to mails and attachments in the database.
- the archived e-mails and attachments are only accessible to the employees who sent or received them.
- a focus on storage and reducing the load on the e-mail server: the accent lies on the preservation of the bits of e-mails and attachments, not on the preservation of the conceptual record.
- insufficient long-term readability guarantees: large dependency on a closed or non-transparent database systems, storage in proprietary, non-standardised or closed container computer files, use of compression, no general archiving solution for all types of attachments, etc.

---

<sup>19</sup> See *Advies & Analyse*, Report no. 4, for a thorough discussion of the functionalities, and the advantages and disadvantages of each archiving solution: ANTWERP CITY ARCHIVE, *E-mail archivering*, Advies & Analyse 4, April 2002 (<http://stadsarchief.antwerpen.be> → Toezicht op archivering → Standpunten en rapporten → 4 Emailarchivering). The evaluation of commercial packages started in 2002. Since then the Antwerp city archives has continued to follow the evolution of the market, but has found that the shortcomings of the commercial archiving solutions remain the same.



*No structural solution* Regarding to long-term readability, accessibility and records management (f.i. filing), the commercial packages provide no real added-value compared with the e-mail systems themselves. They are designed mainly to reduce the load on the e-mail servers by managing old e-mails and attachments. For this reason, large (virtual) mailboxes and information isles continue to exist within the organisation. In addition, the various commercial archiving solutions have in common that they require the installation of new hardware and software (e.g. servers, server software, database system, client software), for which large investments in resources and personnel are needed.

*Conclusion* In consultation with Digipolis, the city archives of Antwerp decided not to use a commercial archiving solution and to give priority to developing our own archiving strategy and procedure within the existing MS Exchange and Outlook e-mail configuration. Several other options for adding contextual and transmission data were also investigated, but they offered no added value compared to the proposed DAVID solution.

### **3. FILING E-MAILS AND ELECTRONIC DOCUMENTS**

#### **3.1 Building a classification system and creating electronic files**

*Importance* When starting e-mail archiving, much attention should be given to the design of a good classification system in which all electronic records regardless their provenance or the application with which they are created can be managed. The e-mail archiving procedure provides a good opportunity for the creator to organise his electronic records management in a coherent, structured and organised way. By means of the electronic filing system, structure can be given to the way electronic documents are managed and kept. Doing so, the electronic filing system becomes the information asset of the organisation. The success of the e-mail archiving procedure will be depending on the user friendliness of the classification or filing system. The e-mail user will only add e-mails and attachments to electronic files if he easily knows where to file the documents and if he can also find them quickly afterwards. Measures such as limiting the maximum mailbox size will only encourage the user to archive if he can easily find his way in the folder structure. Otherwise this will lead to storage in personal mailboxes or off-line folders, and to not approved disposals.

*Setting up an electronic filing system* In consultation with the archival service, the agency creates the shared folder structure within which electronic records are filed. The folder structure is the product of a consultation group that is specially constituted for this purpose. In addition to the superintendent archivist, this consultation group includes the contact person of the archival service in the agency, the LAN manager and the administrative employees who have a mandate or responsibility in this area. The objective of this consultation group is to create a logical and well-organised electronic filing system. One can develop a good filing system for all electronic office documents by following various paths. The paper or existing electronic filing system might serve as a basis. If there is a well-functioning paper filing system in the organisation, the folder structure can be adapted to it. Another possibility is a thorough investigation and revision of an existing electronic filing system. If the

creator does not have a paper or electronic filing system, one must start from scratch.

*Digital ArchiVing: guldeline & aDvice, nr. 3* In a guideline for electronic records management, the DAVID project has established general rules and recommendations for the development of classification systems (Digital ArchiVing: guldeline & aDvice, no. 3)<sup>20</sup> so the central folder structure can accomplish the intended objectives, namely: electronic file creation, indication of the context, and sharing of information. The most important basic principles and rules are:

- construct a logical and well-organised classification structure. Be sure that users clearly know in which folders they have to save documents in and how they can find them again afterwards.
- base the classification structure on the workprocesses (tasks and activities) of the creator
- build the structure up from the general to the particular, first internal tasks and then external tasks
- correlate the classification structure with the paper filing system
- include a structured filing code as the first part of the folder name. Possibly adopt the filing code of the paper files. Think carefully about a structured rubrication, and about the composition and structure of the filing code. Also assign filing codes to the subfolders.
- keep the number of levels under control: limit it to about five levels
- give the folders a semantic and process-related folder name. Do not reuse any folder names for subfolders.
- take the limitations of the ISO-9660 standard into consideration. Complying with this standard is not only important when using CD's as a transfer or archiving medium, it also ensures that hyperlinks to internal documents can be forwarded rather than having to forward documents as attachments each time. The main points for consideration are:
  - assign folder names of maximum 31 characters
  - do not use spaces but underlines, or write words together as one word
  - only use the characters: A-Z, 0-9, \_
- make fixed agreements for the use of abbreviations. Document the abbreviations that are used.

*Platforms for the classification system* The classification system can be hosted by various IT infrastructures. A classification structure can be constructed in the file system of regular operating systems or can be stored in a records management application. File systems of operating systems have the advantage that they are present everywhere and that the every user is familiar with their operation and the associated management software. Their disadvantage is that they are designed for the management of computer files in general, and not for electronic documents in particular. Specific records management functionalities are lacking in the common tools with which computer files are managed (Windows Explorer, Linux Nautilus File Manager, Apple Finder, etc.). Version management, registering metadata at series or file level, access control, advanced search possibilities, etc. are the specific functionalities of records management applications.

For electronic records management, the city of Antwerp decided to develop their electronic filing systems on shared file servers. Not only is the number of city agencies with a records management application limited, the introduction of an

---

<sup>20</sup> This guideline is an application of Digital ArchiVing: guldeline & aDvice, no. 3 (<http://www.expertisecentrumdavid.be/davidproject/teksten/guideline3.pdf>)

electronic classification system is an important change in records management. The workprocess-based filing of electronic documents in a hierarchical structure of series, dossiers and folders, is for many users a new way of managing of their electronic documents. Many use their own classification system (per year, per document type, etc.) or a personal method for the assignment of computer file names. For this reason, the implementation of records management applications sometimes fails. Since the basic principles of an electronic filing structure are the same for a computer file system as for a records management application, it was decided to first familiarise the user with the new operating procedure for electronic records management within the existing IT environment<sup>21</sup>.

This step-by-step approach also has the advantage that the desired functionalities for a records management application gradually become clear. This gives the users, the records manager and the archivist a better insight into the added-value that a records management application can provide, so a more targeted search can be made for a suitable product on the market.

*Maintaining quality*

It is recommended to provide some form of quality control, so the classification structure remains well-organised. To this end, one or more people can be made responsible for each classification system or agency. It is also best if these people supervise the rubrication of the filing codes.

## **3.2 Registering metadata**

### **3.2.1. ABOUT SERIES AND DOSSIERS**

*Descriptive and administrative metadata*

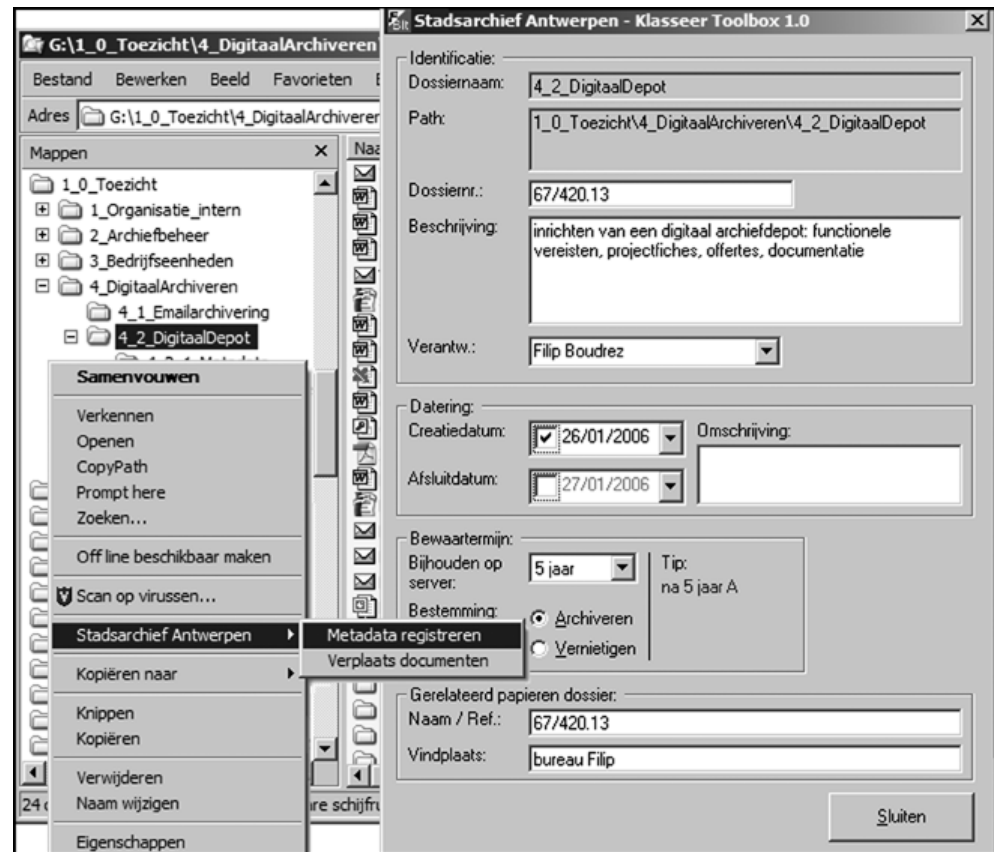
In addition to several items of descriptive metadata, it is also advisable to include some administrative metadata about the series and the dossiers. The name of the process owner or the records manager, the administrative retention period and the final destination are examples of this. The registration of such metadata is usually one of the standard functionalities of a records management application. If an electronic classification system is built into a file system of a regular operating system, these metadata can be kept in a separate document.

A compromise was chosen in the implementation for the administration of the city of Antwerp. Records management applications are not present in every agency, whereas regular operating systems are. Therefore the decision was made to build an extra customisations within a regular operating system. In spite of the limitations of a regular computer file system as a storage place, it is still possible to register metadata about the series and the files. With the help of an ad hoc tool that was developed, metadata are added to a selected folder. These metadata are stored in a XML metadata document and are kept in the folder to which they relate. This XML

---

<sup>21</sup> Recent developments in various document management systems and records management applications make it possible for documents to be found quickly even though they are not organised in a folder structure. Finding documents then occurs mainly on the basis of indexes and by searching on designated metadata (the content description, for example). Several organisations experimented with this operating procedure, but have in the meantime returned to the system of a folder structure: the assignment of descriptive metadata at 'check-in' requires a certain amount of time, employees are accustomed to arranging documents in folders so contexts are clear, it is not always easy to find documents quickly on the basis of metadata or a full-text search, etc.

metadata document is given the attributes of a hidden system file so the metadata are only editable by a custom interface (shell extension of the Windows explorer). It is not intended that every user is supposed to assign metadata for series and files. This task should be performed by the civil servant responsible for records management within his agency.



**Illustration 3:** With the help of this tool, metadata are assigned to series and files automatically and manually.

*Relationship among related documents*

The export of e-mails and attachments to a central electronic filing system leads to the creation of electronic files that contain the electronic records. This centralizes all electronic records of the organisation. In addition to the electronic documents, the organisation will, in many cases, also have paper records for the same series or files. The paper and electronic documents should be placed in a relationship with each other by harmonising the electronic classification structure with the paper filing system, and if possible by using the same filing or registration codes for the paper and for the electronic series and files. On the basis of this shared filing or registration code, the paper and electronic items can be retrieved relatively fast. In both folders a reference can also be made to the related paper or electronic file. One simply places a reference note in the paper dossier. In the metadata of the electronic file, the number and/or the location of the related paper dossier can be indicated.

### 3.2.2. ABOUT E-MAILS

- The need for 'capture'* The essential transmission data about the sent and received e-mail messages are present in the e-mail system. But these data are not always saved or presented to the user in a static or structured manner. This is the case, for example, with the date and time of receipt of a received e-mail. When adding e-mails to electronic files, these data are not always brought outside the e-mail system and linked to the e-mail message in a persistent way. Because of this, the risk exists that they will be changed or lost. The registration of the essential transmission data, and linking them in a persistent way to the e-mail messages, are therefore important points for consideration when filing e-mails.
- Metadata to be registered explicitly?* For e-mails with record status, the following metadata are explicitly registered:
- the e-mail address of the sender
  - the name and the e-mail address of the authorised delegate
  - the date and the time of sending
  - the date and the time of receipt
  - a reference to the filed attachment(s)
  - a reference of the archival context within which the e-mail message is situated
- The other essential transmission metadata can be retrieved without difficulty for filed e-mails without one having to pay any attention to this when filing, and without needing the e-mail server to retrieve them.
- Capture moment* Ideally, transmission metadata should be registered as soon as possible after the time of sending or receipt. Otherwise the possibility increases that these data will no longer be accurate. In any case, at the very latest, these metadata must be registered at the time of filing. From a technical point of view it is absolutely essential to register the e-mail address of the sender and possibly of the authorised delegate.
- E-mail addresses of the sender and the authorised delegate* With the standard security-settings, both e-mail addresses are protected against viruses or other malafide computer programmes that want to use these data to propagate themselves<sup>22</sup>. The Outlook object model provides a 'SenderName' attribute of the object 'Mail item', but it does not necessarily return the e-mail address of the sender<sup>23</sup>. As long as an e-mail is preserved within the MS Exchange and Outlook environment, one can gain access to the e-mail address of the sender and the authorised delegate in one way or another. With filed or exported e-mails, however, this is not necessarily possible. Since the link between these e-mails and MS Exchange is broken when they are exported, the e-mail address of the sender or his authorised representative is no longer retrievable via the server (for example, via CDO<sup>24</sup>).

<sup>22</sup> For this reason, in the object model of Outlook 2000 and 2002 the e-mail address of the sender is not provided as an attribute of a mail item. In the object model of Outlook 2003, the "Mailitem.SenderEmailaddress" attribute is present but this code is only implemented if the plug-in is set as a trusted code.

<sup>23</sup> The attribute 'SenderName' returns the first text value of the display name of the sender. For an Exchange user this is usually the surname and first name of the sender. For other users this can be the name and first name, the SMTP e-mail address or a combination of both.

<sup>24</sup> CDO (Collaboration Data Objects) is an alternative method of dealing with Exchange server and Outlook data. For use on the client side, CDO 1.21 must be installed as a part of MS Outlook.

*Capture and storage place* Since all transmission metadata are known by the e-mail system, they can in principle be captured completely automatically. For the contextual metadata, the intervention of the sender or the addressee is required. An obvious and safe storage place for these data is in the filed e-mail itself. By embedding the essential metadata, they remain permanently linked to the e-mail message to which they are related. This does not have to occur for each e-mail document, but only for the messages with record status.

### 3.3 Filing e-mails and attachments

*Export from e-mailsystem and import in classification schema* In the DAVID strategy for the archiving of e-mails, both e-mails and attachments are filed in the series or files to which they are related. This action involves moving e-mails and attachments from the e-mail system to the place where the electronic classification system is hosted. If storage is done in a common computer file system, the e-mails and attachments must simply be exported to the series or the file to which they belong. When a records management application is used, the e-mails and attachments not only must be exported, but they must also be checked in immediately. In the latter case, ideally the e-mail software and the records management application should be integrated, so e-mails and attachments are placed in the electronic classification system in an efficient and automated manner (Moreq: 6.4.1; 11.1.13)<sup>25</sup>.

*When to file?* Ideally, e-mails and attachments with record status should be filed as soon as possible after receipt or sending. Important arguments for this are:

- accuracy of the metadata
- good electronic file creation: as long as e-mails and attachments with record status are not ingested in the electronic classification system, they are actually not yet captured as records of the organisation
- consultability by third parties / colleagues: filed e-mails and attachments can be shared with colleagues
- safety: storage in the electronic classification system is safer than in the e-mail system

In practice, however, it is also possible to preserve e-mails and attachments with record status in the e-mail system. An e-mail user can keep his e-mails in his 'IN BOX' or 'SENT ITEMS' or can build a folder structure (for example, in his 'IN BOX' or in a .pst file). Most e-mail client programmes provide several functionalities for organising and searching through received and sent e-mails. Although, for the above-mentioned reasons, this is not the most desirable situation, it cannot be avoided in practice. From a records management and record-keeping point of view it is, however, important to point out that preservation in the e-mail system may only be temporary at the most.

Regardless of the time at which e-mails and attachments are filed (immediately after receipt or sending, or after temporary preservation in the e-mail system), the same requirements apply for the filing process.

In addition to the registration of the essential metadata, another important point for consideration when exporting is the computer file format in which the e-mails are

---

<sup>25</sup> For such a functionality, integration between the e-mail server and the DMS/RMA will be needed in most cases.

*File formats for  
filed mails and  
attachments*

saved. Most e-mail client programmes support several export formats (.eml, .txt, .html, .msg, .oft, .rtf, etc.). Criteria for selecting an export format are:

- the inclusion of all essential elements of the e-mail message
- the embedding of the transmission and contextual metadata must be configurable
- the reading, answering and forwarding of the filed e-mail must remain possible after exporting and filing
- it must be a suitable source format for migration to the preservation format for e-mail.

It is advisable to establish one export format for filed e-mails in the organisation. Ideally, this export format should also be the archiving format for e-mails, but in practice the suitable archiving formats for structured text documents (PDF/A, XML, ODT, TIFF) cannot be reopened, answered or forwarded by the e-mail client programme without difficulty. For these reasons, the Antwerp city archive chose the message format (.msg) as the export format for filed e-mails. The message format is not a suitable archiving format, but is the undocumented and native application file format of MS Outlook. The filed e-mails can be reopened in MS Outlook, so they can still be read, answered or forwarded<sup>26</sup>. This is an important condition for getting e-mails with record status filed as soon as possible after receipt or sending. If this is not possible, or if the e-mail client does not permit it for the export format, then many users will tend to keep their e-mails in the e-mail system for a long time and postpone filing them in the electronic classification system. The selection of .msg as the filing format means that e-mails with archival value still must be converted to a suitable archiving format before they are included in the digital repository<sup>27</sup>. The attachments are exported in their original computer file format and in most cases will also have to be migrated to a suitable preservation format.

*Separating e-  
mails and  
attachments*

Before moving e-mails out of the e-mail system, the attachments are exported and separated from the e-mails. For long-term preservation, it is better for e-mails and attachments to be separated. Since they are separate documents and are only related to each other, it is not good for them to be preserved as one computer file. By preserving them separately, the documents can be identified and reused more easily. It's also very likely that the various types of electronic documents (text, illustrations, audio, video) will require different approaches for tackling the digital durability problem. Separating the attachments from the e-mails makes it possible to use the most suitable archiving solution for each type of electronic record. However, within the default configuration of MS Outlook it's possible to export e-mails with embedded attachments to a file system.

*Filing with the  
standard Outlook  
functionalities*

For exporting e-mails and attachments from the e-mail system to an electronic classification system build in a regular computer file system, it is possible to use the standard functionalities of MS Outlook. A pilot project for e-mail archiving in the agency for human resources of the city of Antwerp, however, indicated that this is not easy within the standard configuration of MS Outlook:

- a lot of users did not pay no attention to the file format in which e-mails are saved. E-mails were saved as .msg, .txt, .rtf, .html and .oft files<sup>28</sup>.

<sup>26</sup> See also DOD 5015-2, C2.2.6.8.8.

<sup>27</sup> A variation on this approach is the simultaneous exportation of all e-mails in an export format that is also a suitable archiving format. Although this is perfectly implementable technically, this option was not retained for implementation by the city of Antwerp. Only filed e-mails with archival value are migrated to the suitable archiving format after selection.

<sup>28</sup> In MS Outlook the formatting of the message body determines which file format is preselected as the export format.

- attachments with record status were often not filed (for example, with .txt or .html as the export format) or they were embedded in the exported e-mail message (for example, with .msg) which causes them not to be easily findable and reusable as separate documents.
- separating e-mails and attachments, and recording the mutual relationship is labour-intensive when this must be done manually. Furthermore, the chance of making errors is quite great.
- the manual registration of metadata is experienced as being too time-consuming and was therefore insufficiently applied.

### 3.4 Customisations

*The need for customisation* The default configuration of MS Exchange/Outlook provides no specific functionalities for the registration of all essential metadata or for the user-friendly filing of e-mails in accordance with administrative and archival needs. Thus it was necessary to develop a purpose-built customisation.

*Functionalities* Desired functionalities for this customisation are:

- the registration of all metadata that are essential for records management and record-keeping purposes
- the preservation of these metadata in a static, structured and reusable manner
- the linking of the e-mail message and its metadata in a persistent and unbreakable way
- the user-friendly export of e-mails and attachments in which:
  - the required user interaction is kept to a minimum
  - the making of (human) errors is avoided as much as possible
  - a selection can be made as to which attachments are filed or not
  - the file names of the filed attachments can be adapted
- the pre-programming of the export / filing format, in this case the MS Outlook message format
- the separation of the e-mail and the attachments when they are filed
- the indication of the relationship between the e-mail and the associated attachments

*Within MS Exchange/Outlook* A customisation within the MS Exchange/Outlook environment was decided on rather than searching for (new) software that provides the desired functionalities. This offers several advantages. First, e-mail users can assign the contextual metadata to the e-mails themselves. This is important for the sake of metadata quality: the senders or the recipients are familiar with the meaning and the function of the e-mails, and are best placed in the organisation to add context to the messages. Second, registration of the metadata can occur immediately or as soon as possible after sending or receipt, which is important. Retroactive operations are not feasible and will seldom reach the quality of immediate registrations. The third advantage is that most e-mail users are familiar with the MS Outlook mail programme and do not have to learn to work with a completely new application.

*Two elaborated solutions* For the registration of metadata and the user-friendly filing of e-mails and attachments, the Antwerp city archives developed two solutions within the MS Exchange/Outlook environment<sup>29</sup>:

---

<sup>29</sup> A third alternative is a combination of the two solutions: using the adapted form for filing received e-mails and the plug-in for filing sent e-mails. This alternative was worked out in

- customising the e-mail headers
- adding a plug-in to MS Outlook with records management functionalities.

Both solutions provide similar functionalities for the registration of metadata and the filing of individual e-mails, but the technology used for the two solutions is substantially different.

### 3.4.1 CUSTOMISED E-MAIL FORM

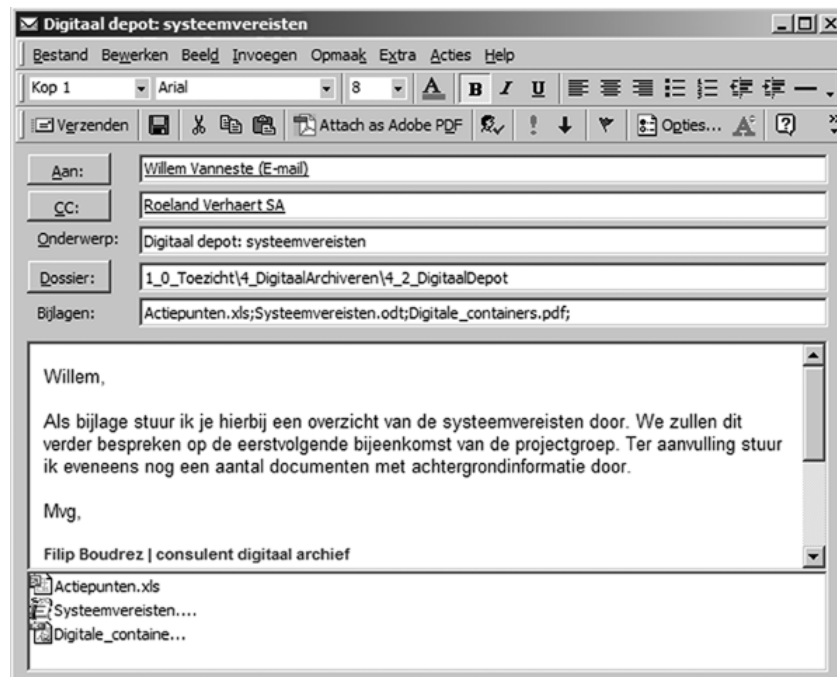
<i>Adapting the default e-mailheader</i>	In this first customisation, the standard e-mail header for received and sent e-mails is replaced by an adapted e-mail header. The standard e-mail form was customised with additional controls and fields <sup>30</sup> . Both the e-mail header of the composition page and of the reading page are adapted, as it must be possible for both the sender and the recipient to add metadata to the message and to file e-mails.
<i>A scalable solution</i>	Working with an adapted e-mail form is a scalable solution, an important point to consider when implementing an archiving solution in a large organisation. The adapted e-mail form can be made available to each e-mail user centrally from the e-mail server. The form only has to be published once in the central form library of the Exchange server. The change involved in adapting the e-mail form only has to be made once. At the client level, only the windows registry has to be modified so the adapted form is automatically displayed when a user composes a new mail or opens a received mail. For this, Outlook 2000 or later is required. This modification of the Windows registry has to be done only once and can be done automatically when logging on to the server. This solution can also be applied in a webmail environment <sup>31</sup> .
<i>Transmission metadata</i>	For the registration of the transmission metadata 'date and time of sending' and 'date and time of receipt', the reading page is expanded with these fields so these metadata are part of the e-mail itself in an explicit and static way. Both items of information appear in the header of a filed e-mail (in contrast with the standard e-mail form). Since these items of information are present in the e-mail system for each mail and can be retrieved automatically, the e-mail user does not have to do anything for this at all. It is done for every e-mail, also for e-mails without record status or archival value.
<i>Contextual metadata</i>	To know the archival context of an e-mail, one must know the work process and other records to which it is related. To this end, both the mail composition page and the mail reading page are expanded so both the sender and the recipient can add these data to the e-mail. In the composition page, a textbox is provided for the registration of a filing code ('DOSSIER') and the file names of the attachments ('ATTACHMENTS'). These same fields are also provided on the reading page. In addition, on the reading page an extra textbox is provided for the classification or registration reference of the recipient ('DOSSIER ADRESSEE').

---

practice, but was not tested extensively.

<sup>30</sup> Additional control elements alone are insufficient: control elements serve only for the display of data and not for storage. The information is saved in fields. Without these fields, after closing or sending, the content of the control elements is lost.

<sup>31</sup> By means of Javascript embedded in the HTML page, the adapted e-mail form can be retrieved from the mail server.



**Illustration 3:** Customised e-mailheader for composing e-mails, with the extra field 'CASE FILE' [Dossier] and 'ATTACHMENTS' [Bijlagen]

The 'ATTACHMENTS' field

In the adapted e-mail header, an additional field is provided for the registration of the file names of the attachments. Since Outlook 2002, such a field is default part of the e-mail header when attachments are added or present in a received e-mail. Still it is advisable, also in Outlook 2002 and 2003, to provide a separate field for the file names of the attachments. The field that Outlook 2002 and 2003 automatically adds is dynamic in nature. This means that the file names of the attachments disappear when the e-mail and the attachments are separated from each other during filing. The archival bond among these related documents would in this way get lost and would no longer be reconstructible.

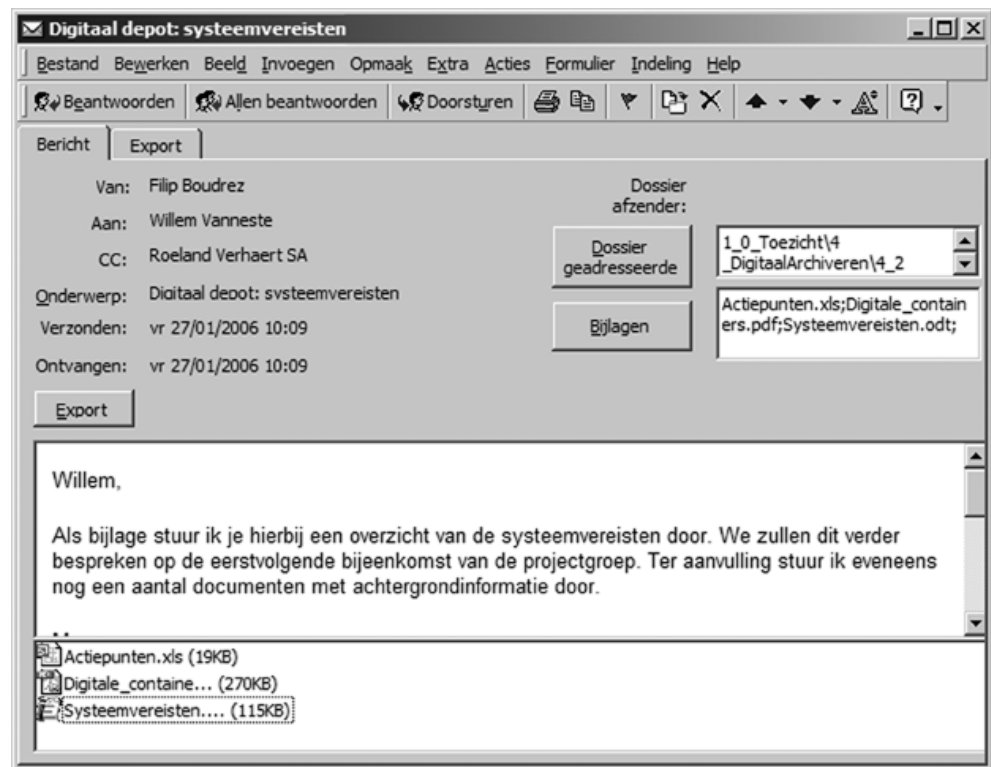
Just like the transmission metadata, the file names of the attachments can be captured completely automatically. With the help of a Visual Basic script, the additional header field 'ATTACHMENTS' can be filled in automatically. VBScript is a 'light' version of the programming language, Visual Basic for Applications (VBA), and can be linked to an e-mail form<sup>32</sup>. Since VB scripts can be included in HTML pages, this solution is also applicable for webmail<sup>33</sup>. When the user adds an attachment by dragging or pasting it, the 'ATTACHMENTS' text field is filled in automatically on the composition page. When a received e-mail is opened, the 'ATTACHMENTS' text field on the reading page is also filled in automatically. This is interesting because e-mail users from outside one's own organisation do not have access to the adapted e-mail forms. Filling in or adapting the information manually remains possible, however.

<sup>32</sup> The addition of scripts is not a problem because the central form library is automatically viewed as a trusted environment. The warning message for possible macro viruses is therefore not shown.

<sup>33</sup> Not only VBScript, but also Javascript, Java Applets and ActiveX elements can be linked to HTML pages.

*Classification or registration reference*

The assignment of a classification or registration reference cannot occur completely automatically, however. For this, the intervention of the user is required. The civil servant indicates the electronic series or files in the classification structure to which the e-mail belongs. For looking up and retrieving the corresponding folder name, a VB script can be used in combination with a common dialog, so the sender or the recipient only has to browse through the classification structure and to select the appropriate folder name. In the e-mail header, the folder name and the names of the two parent folders are shown. The complete path of the selected folder is written to a hidden text field (see below). Whether the sender or addressee actually assigns a filing or registration code will depend to a large degree on the filing or archiving reflex. The retrieval of folder names must become a routine action that can be encouraged by training and instruction, but that requires a certain amount of discipline and carefulness.



**Illustration 5:** E-mail header for received / sent e-mails with the added fields 'DOSSIER SENDER', 'DOSSIER ADDRESSEE', 'ATTACHMENTS', 'SENT' and 'RECEIVED'.

*Storage place for metadata*

The transmission and contextual metadata are saved in the filed e-mail message itself. These data are preserved in visible and some hidden fields in the e-mail header. The user can still edit most of the metadata if needed.

*Filing attachments*

The e-mail form also provides some functionality for the (semi-)automated filing of the attachments of an e-mail. When filing e-mails with attachments, it is better to save the e-mail message and the attachments in the electronic folder as separate digital objects. If an e-mail contains one or more attachments, a second tabpage appears in the opened e-mail in which the file names of the attachments are listed. The user can indicate by checking or unchecking the check boxes which attachments will be filed and whether they will be filed together with the e-mail

message in the same folder or not. If necessary, the user can change the file name of the attachment so it is meaningful. The relationship between the e-mail and the attachments is indicated by registering the file names in the designated field in the e-mail header. In this field, only the (adapted) file names of the filed attachments are shown.



**Illustration 6:** The end user can select which attachments to file and can change the filenames of the attachments.

#### *Deleting e-mails in MS Outlook*

After exporting an e-mail, the e-mail usually remains in the e-mail system. In principle, this representation of the e-mail in MS Outlook may be deleted. With the adapted form, after filing e-mails and attachments, the user gets the option of deleting or retaining the e-mails in MS Outlook. Ideally, the e-mails in MS Outlook should be deleted after filing as much as possible to reclaim space in the mailbox. These e-mails are then placed in the folder 'DELETED ITEMS' so they can still be recuperated if needed. If the user decides to keep an e-mail in his mailbox, that e-mail message is automatically given the status 'FILED'. This can prevent the same e-mail from being filed a second time, and the user can quickly select all filed e-mails in his mailbox and delete them.

#### *Defining an e-mail document model*

By adapting the e-mail form, the archivist has the opportunity to define the document model for e-mails in his organisation. This gives the archivist the chance to think carefully about the data fields and the (internal) structure of e-mails in advance, and to define the relationships among the various components. By doing so, the appraisal and the needs for long-term preservation can already be taken into consideration. One can, for example, develop the document model round the essential components of e-mails. The internal structure of the record can be archived more easily if the e-mail is well-structured from creation on.

### 3.4.2 A PLUG-IN WITH RECORDS MANAGEMENT FUNCTIONALITIES

The second customisation adds several new functionalities to MS Outlook. They are built into the e-mail client itself. When MS Outlook starts up, these extensions are automatically loaded so they are available for the recipients or addressees. After installation of the plug-in, the menu and the standard toolbar in MS Outlook are expanded respectively with an 'ARCHIVE' item and a 'FILING' button. This last button

also appears in each Outlook window for received e-mails. Nothing is changed on the e-mail form: the end user goes on working with the standard e-mail headers.



**Illustration 7:** The customisations to MS Outlook. A 'FILE'-button [klasseer] is added to the main toolbar and a 'ARCHIVE'-item [archief] is added to the menu bar. One or multiple e-mails can be filed with the 'FILE'-button. With the options in the 'ARCHIVE'-item, one can file the complete contents of an Outlook-folder or one can archive appointments from the Outlook calendar.

*A scalable solution?* A big difference with the customised e-mail headers is the installation process of this option: the plug-in must be installed on each client computer. For small or medium-sized creators, the installation can be done manually. For large organisations, a method of automatic distribution and/or pre-installation by means of preps/ghosts will be more appropriate. For installation on Windows XP operating systems, one must have administrator rights (for the installation of system dll's and the modification of the Windows registry).

*Registering metadata* When e-mails are filed, the plug-in registers the same transmission and contextual metadata about the e-mail messages as the customised e-mail form. Here the necessary transmission metadata are registered completely automatically. One important difference with the e-mail form is that the plug-in has several options for obtaining the e-mail address of the sender and of his delegate. When a first attempt does not result in a valid e-mail address, there are still at least two back-up procedures which can be performed by the plug-in.

With regard to contextual metadata, the (adapted) file names of the filed attachments are also automatically captured. For the destination folder, the user must enter the appropriate dossier or folder just like he does with the e-mail form. For this he can use a browse function so only the relevant series or file name must be retrieved from the classification system. The plug-in remembers the last ten selected target folders, which in many cases enables the end user to quickly make the appropriate choice.

*Storage place for the metadata* The transmission metadata and the contextual metadata are saved in the filed e-mail message itself. These data are preserved in self-defined user properties in the e-mail message. The embedded metadata are not visible and cannot be edited by users with average PC skills.

*Filing several e-mails at the same time*

In contrast with the customised e-mail form, the action range of a MS Outlook plug-in is not limited to one e-mail. Multiple e-mails can be filed at the same time. A user can select several mails and add them to a series or file in one operation, or he can even file the complete content of one selected Outlook folder (including subfolders). This last option is especially interesting for the retroactive filing of e-mails and attachments that were kept in the e-mail system for a while.

*Filing attachments*

Just like the e-mail form, the plug-in provides several functionalities for filing e-mails and attachments as separate electronic documents in the same series or file. When filing individual e-mails, in more or less the same way as with the adapted e-mail form, the user can decide which attachments will be filed or not, and the file name can be adapted if desired. When filing several e-mails at the same time, all attachments are filed with their existing file names.



**Illustration 8:** The end user can select which attachments to file and can adapt the file names.

The relationship between the e-mail and the attachments is preserved by embedding the file names of the filed attachments as metadata in the e-mail. These metadata are not visible to the end user, however. To visibly indicate the mutual relationship, the filed attachments are replaced in the e-mail message by shortcuts to the corresponding documents in the same folder<sup>34</sup>.

*Deleting e-mails in MS Outlook*

At the end of the filing process, the user is asked whether the e-mail may be deleted in MS Outlook. If the end user answers 'no', the e-mail is given the status 'FILED' in MS Outlook. The e-mail in MS Outlook contains all attachments that were sent, thus also the attachments that have not (yet) been filed. This makes it possible for e-mails and/or attachments to be filed in different folders.

<sup>34</sup> The opening of shortcuts also introduces a security issue. With the standard security settings, when a user opens a shortcut in an e-mail he first sees a warning window. One can avoid this by setting a low security level for attachments (Outlook) and by deleting the \*.lnk extension from the designated file types (Windows). All of this can be automated (low security level for attachments: modify Windows registry of client PC's; delete \*.lnk extension from designated file types: define as part of a group policy) or can be set individually for each client PC.

### 3.4.3 A COMPARISON OF THE TWO CUSTOMISATIONS

Within the MS Exchange/Outlook environment, the Antwerp city archives developed two solutions for the registration of metadata and for the user-friendly filing of e-mails. Both alternatives have specific advantages and disadvantages that are compared in the following table.

	<i>E-MAIL FORM</i>	<i>PLUG-IN</i>
<b>METADATA:</b>		
registration:	transmission metadata: automatic file names for attachments: automatic	transmission metadata: automatic file names for attachments: automatic
context metadata:	browse in electronic classification system	browse in electronic classification system
registering e-mail address of sender:	limited possibilities	several alternatives
time:	registration immediately on receipt/ sending or when filing	registration only at the time of filing
storage method:	embedding (additional fields in the e-mail header)	embedding (self-defined user properties)
visible for end user:	partially	no
<b>FILING:</b>		
reflex:	standard provision: additional header fields act like visual reminders and encourage filing	additional mechanisms needed to encourage users to file
number of items:	only individual e-mails	possible for individual e-mails, selected e-mails, or the complete content of one Outlook folder with the option of including subfolders
retroactive:	not practical	can be provided
checking of computer file names:	filters out disallowed characters	filters out disallowed characters
<b>TECHNICAL:</b>		
platform:	MS Exchange, Outlook 2000/2002/2003, MS Internet Explorer (5.0 and later)	Outlook 2000/2002/2003
installation:	server: publish form client PC's: modify Windows registry, install OCX component <sup>35</sup>	server: define security settings client PC's: install plug-in, modify Windows registry
robustness:	only limited error-handling is possible	extensive error-handling is possible
standard Outlook and Windows security:	no problems	triggers certain 'warnings' that can be avoided by various workarounds
integration in webmail:	possible	not possible

<sup>35</sup> This OCX component is used during automated browsing through the electronic classification system and is a default installed with certain versions of MS Office. During the de-installation of software that uses this same component, it might be deleted.

---

Outlook quirks: a few Outlook functionalities are no longer available

Outlook does not always shut down correctly, which causes the plug-in not to be loaded when (re)starting

Both possibilities for filing individual e-mails were put in practice and compared with each other. The Antwerp city archives was responsible for the development of both the customised e-mail form and the plug-in. Digipolis, the information-technology partner of the city of Antwerp, investigated both alternatives in the technical area. The technical research did not provide any specific arguments for or against either of the two possibilities. On the basis of a functionality comparison by users, a decision was finally made in favour of the plug-in. The plug-in was experienced as more user-friendly by most of the testers.

### 3.4.4 THE FILING TOOLBOX 1.0

*Encouraging to file*

After the comparative technical and user research, the plug-in for filing individual e-mails was refined further. First, several additional mechanisms were added to the plug-in to encourage the e-mail user to file e-mails with record status. This was found to be necessary because the plug-in itself does not remind the user in any way about the need for filing e-mails and attachments. These additional mechanisms are:

- a warning dialog on loading MS Outlook when the total number of e-mails in the 'IN BOX' and 'SENT ITEMS' folders is higher than a predetermined critical value<sup>36</sup>.
- a query for a destination after an e-mail is closed or sent. When a user closes a read e-mail without filing or deleting it, he is asked to assign a destination to the e-mail message. The options are 'FILE', 'DELETE' and 'RETAIN IN OUTLOOK'. The same question is asked when the end user sends an e-mail. This prevents the folder 'SENT ITEMS' from getting full and glutting the mailbox.

*Filing several mails at the same time*

Through a second adaptation of the filing plug-in, functionalities were added for filing several mails at the same time. This enables the user to:

- select several mails in the same Outlook folder. The selected e-mails and their attachments are exported to the same target series / file.
- select one Outlook folder. The complete content of this folder is filed in the same series or file. The export of the content of subfolders is optional, but if this option is chosen, the subfolders are replicated in the targetfolder.

This last functionality is mainly intended for filing e-mails and attachments that are kept at various places in the e-mail system. This instrument can be used when e-mails and attachments are saved temporarily in the e-mail system or when old mails and attachments have to be filed retroactively. A manual clean-up and archiving procedure would take too much time.

*Assigned file names when filing multiple e-mails at once*

When several mails are filed at the same time, the user is not asked for a file name for each individual e-mail. In this case the file names are assigned automatically. The end user sets which header data are used to compose the file name for the exported e-mail message. These data are:

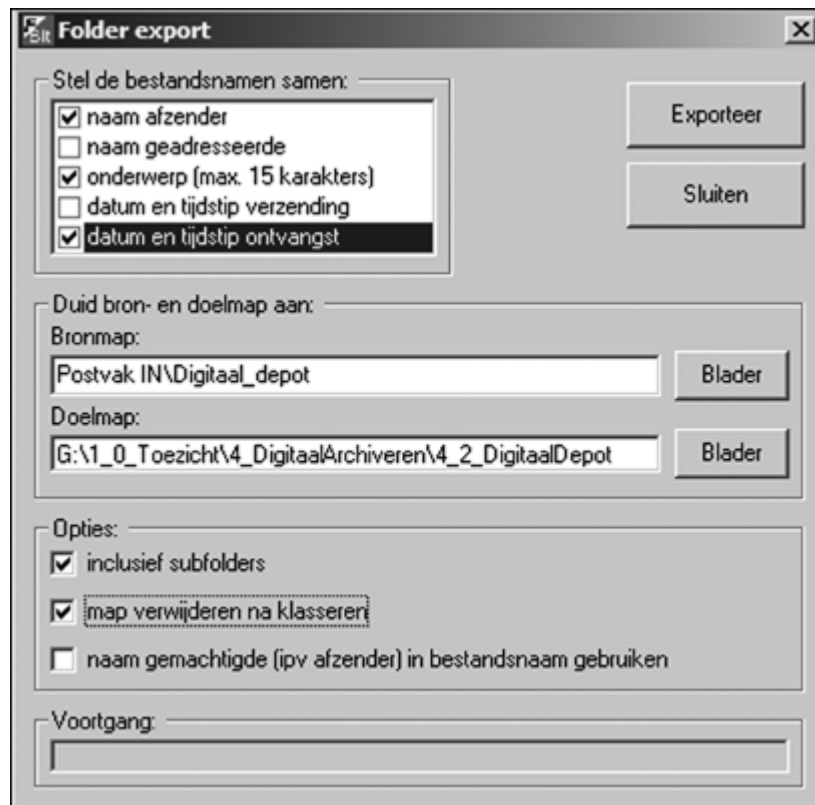
- the name of the sender (possibly to be replaced by the name of the authorised delegate)

---

<sup>36</sup> This value was set at 250 items for the city of Antwerp.

- the name of the recipient
- the subject of the e-mail message (max. 15 characters)
- the date and the time of sending
- the date and the time of receipt.

When several mails are filed simultaneously, all attachments of the selected e-mails are filed under their existing file names. As when filing individual e-mails, the file names of the filed attachments are embedded in the e-mail message as metadata. The attachments themselves are saved as separate documents and are replaced in the e-mail message by shortcuts.



**Illustration 9:** Filing the complete contents of an MS Outlook-folder: all e-mails and attachments are added to the selected case file folder in the classification schema. The end user selects the structure of the file names for the e-mails, while the existing of the attachments will be used.

*Easy retrieval of  
filed e-mails and  
attachments*

The plug-in tool adds some metadata to the filed e-mails and attachments to make retrieval of filed e-mails and attachments easy and fast. The initial goal of this functionality is to mimic the search behavior of MS Outlook in Windows explorer, so users can search their e-mails and attachments in more or less the same way. For e-mails, the name of the user who filed the e-mail and the full subject are registered as file attributes / properties which are accessible and sortable in the Windows explorer. The same counts for the system time of the filed e-mail. Without this plug-in functionality, the filed e-mail would have the date and time of the moment of filing. By adapting the system time, the e-mail has the date and time of receipt. Attachments of the MS Office suite, get in the comments field a reference to the e-mail they were part of (if the e-mail has been filed). By doing so, there's a cross

reference between filed e-mail and attachment so the archival bond between both records is firmly established.

*Archiving  
distribution lists*

An important point of attention is the use of distribution lists in the organisation and by the users. When using distribution lists, in the e-mailheader only the name and e-mailaddress of the distribution lists is mentioned. To verify which users exactly did receive the e-mail, one has to look up the address book or the contacts list. As this is important data about the e-mail, it's very advisable to capture the members of a distribution lists. The safest method would be implementing this functionality within the normal filing process for every e-mail, and capture and embed this metadata. Although this is technically perfect possible, some tests pointed out that this extra functionality decreases the performance of the plug-in. As alternative, the city archives opted for a periodic capture of the data about all distribution lists available on the e-mail server.

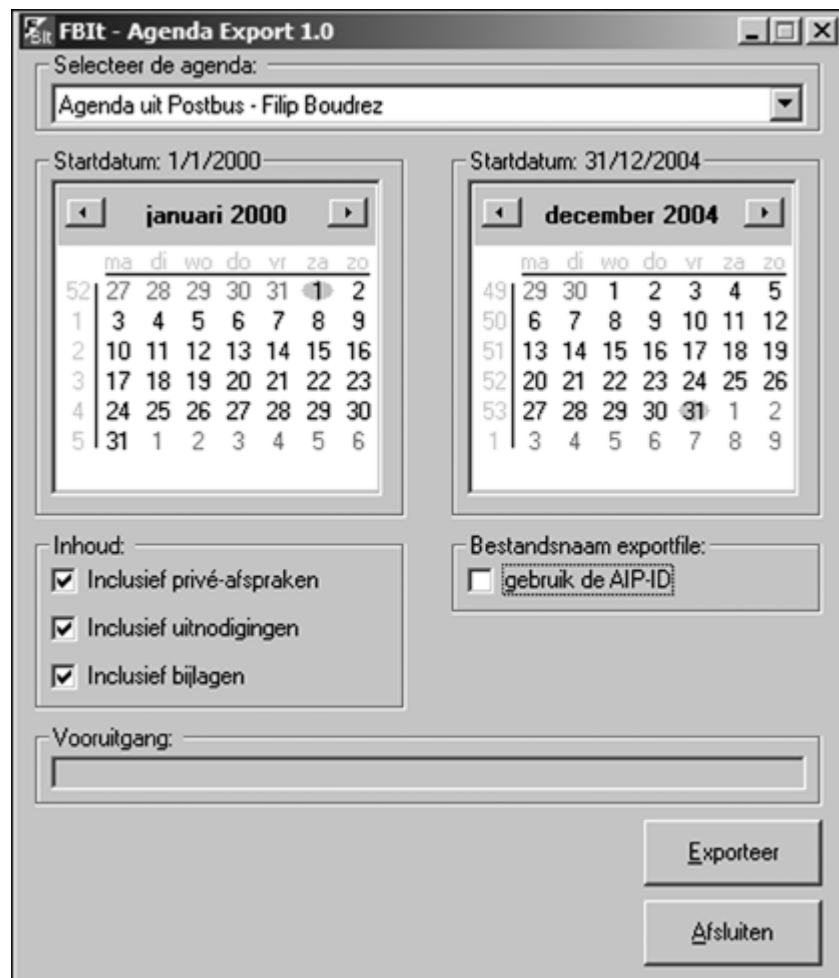
*Archiving MS  
Outlook  
appointments*

Finally, a functionality was added for archiving calendar appointments. As time goes by, appointments in the Outlook calendar occupy a significant amount of the available space in the users mailbox. After archiving the appointments of a certain time span, one can delete them in his calendar and new space can made free in the mailbox.

With this added functionality, the user only has to enter a starting and ending date. He can also decide whether to archive private appointments, invitations for meetings, and attachments. The calendar appointments for the selected period are written to an XML document. This XML document is constructed according to the Expertisecentre DAVID (eDAVID) XML Schema for calendars<sup>37</sup>.

---

<sup>37</sup> See: <http://www.edavid.be/xmlschemas/calendar.xsd>



**Illustration 10:** Archiving appointments of the calendar. The user selects the period for which all appointments (private appointments are optional) will be archived straight into an XML-document.

### 3.5 Implementation

The developed records management procedure is implemented in each part of the organisation using either a project approach or either on a continuous basis. For the latter, the regular training and courses for MS Outlook are extended with a general introduction on e-mail preservation and the tools for filing e-mails and attachments.

In the project approach, the actual implementation is done in phases. First of all, effort is invested on the composition of an electronic classification system. Once this has more or less been brought to a result, training and instruction sessions for the e-mail users are planned. Concurrent with these sessions, the customisations are installed in MS Outlook.

#### 3.5.1 THE ELECTRONIC CLASSIFICATION SYSTEM

During the first phase of the project, work is done to develop an electronic filing system for the part of the organisation where electronic records management is

being introduced. An ad hoc workgroup makes a draft design for the electronic classification system and provides feedback to the users. The archivist serves in an advisory capacity.

*An organisational challenge* Since successful electronic records management stands or falls with a well-organised classification system, it is important to allot the necessary time for this. From a technical point of view, this is the easiest step in electronic records management, but for records management in general, this is the most difficult step. The creation of electronic series and files requires a change in the way most users deal with electronic records and electronic documents in general. That being said, experience also teaches that the planning stage may not drag on endlessly. The ultimate test of the electronic classification system comes when it is put into service. Only after placement into service, it will actually be clear whether the user can find his way around easily when filing and looking up electronic records. This can be monitored, for example, by keeping track of the growth in volume. If this volume does not increase systematically, adjustments or adaptations will be necessary.

Monitoring the quality of the electronic classification system and making adjustments is a continual process. In specific parts of the organisation, people are appointed to be responsible for certain folders.

### 3.5.2 THE TRAINING AND INSTRUCTION OF E-MAIL USERS

*Objective* When the electronic classification system is placed into service, and the filing of e-mails and attachments is first put into practice, it is best to allot the necessary time to the training and instruction of the e-mail users. They continue working within the familiar IT environment (MS Outlook and Windows Explorer), but they need to learn the new functionalities of the Outlook customisations. As they are responsible for the management of their files, learning the basic principles for setting up a good filing system and for good file creation is just as important.

*Training programme* The training and instruction provided by the city of Antwerp consists of three parts. In the first part the users learn which (electronic) documents are records and which are not. The filing of (electronic) documents does, after all, require an effort on the part of the employees, and this effort only needs to be done for documents that belong in the electronic classification system. Next, the basic principles of (electronic) filing are explained: How do you structure the classification system? What is a functional classification? How do you structure series, dossiers and folders? When are files closed or opened? During the third part of the instruction programme, a deeper study is made of e-mail filing and working with the plug-in.

*Points for consideration* A training session usually lasts half a day. The instruction includes:

- outlining the importance of archiving in general and of e-mail archiving in particular: this is important for the motivation and the carefulness of the e-mail user
- teaching the basic principles of filing electronic documents: arrangement of the classification structure, rubrication, assigning folder names and file names
- distinguishing e-mails with record status from e-mails without record status: Which e-mails are preserved? Which e-mails may be deleted immediately?
- functionalities of the plug-in
- filing of e-mails and attachments
- assigning clear and semantic folder and file names

- using e-mail efficiently and composing e-mails that can be easily archived:
  - efficient use of the e-mail system:
    - do not mail internal documents which are available on shared server disks, but only send a link to those documents.
    - fill in the subject field meaningfully.
    - do not add attachments to an e-mail when their content can be included in the message field.
    - do not reply between the lines in the message of the sender.
  - do not send e-mails with an RTF body<sup>38</sup>; use plain text or HTML instead.
  - structure the message by means of white space, and not by means of layout. E-mails do not have a fixed appearance because this is dependent on the client e-mail software used. Not everyone sees the layout.
  - as identification data, insert a signature in the message field of the e-mail body.
  - when using distribution lists: keep an up-to-date copy of the lists of members.
  - keep the printing of e-mails to a minimum. Delete paper copies as much as possible from the paper dossier.

*Assigning  
computer file  
names*

In the folders, electronic documents are identified by a computer file name. The file name indicates which record is saved in the computer file. When exporting e-mails and attachments, one must be careful that the computer files are given unique file names so existing documents will not be overwritten. Digital Archiving: guldeline & aDvice, no.3<sup>39</sup> contains guidelines and recommendations for the assignment of computer file names:

- give the documents a clear and meaningful name. This prevents having to open documents during searches
  - indicate clearly for each document:
    - e-mail: sender/addressee, subject, date (YYYYMMDD)
    - attachments: kind of document, subject, date (YYYYMMDD)
  - if possible include the status or the version number in the computer file name
- do not repeat folder names in the computer file name
- co-ordinate computer file names and titles of documents with each other
- take into consideration the writing of CD's in conformity with the ISO-9660 standard:
  - assign computer file names of maximum 30 characters
  - do not use spaces but underscores or write words together as one word
  - only use the characters: A-Z, 0-9, \_
- retain the original extension of the computer file format in which the document is preserved

### 3.5.3 INSTALLATION OF THE CUSTOMISATION

Concurrent with the training and instruction sessions for the end users, the customisations of MS Outlook are deployed and installed on the client PC's. Ideally,

---

<sup>38</sup> RTF-formatting is a specific feature of MS Outlook. The use of RTF might cause changes in the look and feel of an e-mail as RTF is not always supported by other client programmes than MS Outlook. From a technical and 'filing' perspective it's also not advisable to use RTF formatted bodies as MS Outlook does not expose file handles for pasted images in RTF bodies.

<sup>39</sup> <http://www.expertisecentrumdavid.be/davidproject/teksten/guideline3.pdf>

the users should be able to start working with the new instruments immediately after the instruction session.

Although manual installation of the plug-in by the end user is a possibility, automated installation possibilities were sought for the various parts of the organisation of the city of Antwerp. This can be accomplished by means of an automatic distribution tool or automatic installation via the login script.

## 4. ARCHIVING ELECTRONIC RECORDS

---

The archiving procedure includes: selection of the electronic records with archival value, migration to preservation formats, encapsulation in AIP's, transfer to the repository, and making the information accessible.

### 4.1 Selection of the files with archival value

*The need for selection* To keep the volume of electronic records manageable, the electronic classification system needs to be cleaned out regularly. Organising all electronic records centrally will entail a transfer of electronic documents from the e-mail system to the classification filing system.

*Selection on the basis of records schedules* This selection process is based on the records schedules that are applicable for both paper and electronic records. Usually it will be decided at the series or file level which folders will be deleted or archived after the expiration of their administrative retention period. The actual selection process can occur more or less automatically when preservation periods and destinations are recorded as metadata at the series and file level.

*Moving the folders with archival value* The electronic files without archival value can be deleted, subject to the necessary approvals. This disposition is logged in an XML audit trail of this operation. The electronic files with archival value are extracted from the active electronic classification system. If needed, consultation copies can be left behind (for example, for closed files that are still frequently consulted). These consultation copies should be given the status 'ARCHIVED' and it's recommended to avoid that they are subject of modifications or alterations is best if it is no longer possible for them to be edited. Extraction for archiving involves moving or copying the electronic folders from the active classification system to a location where preparations are made for transfer to the repository.

### 4.2 Archiving metadata

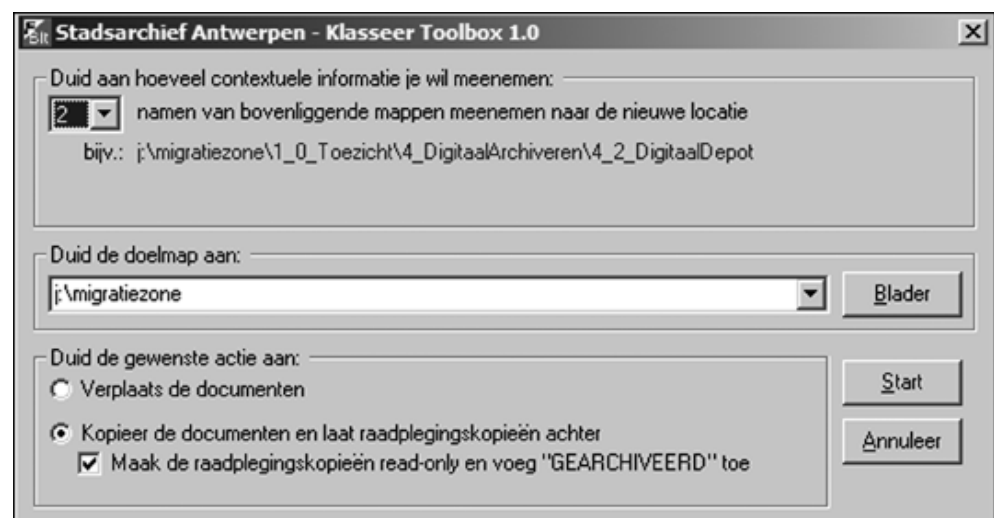
*The need for context information* When the electronic files are taken away from the classification system, it is important that the necessary contextual information is archived as well. The electronic classification structure reflects the context within which series or files are created. Just moving the selected folder is not sufficient for archiving the context as well. The selected folder *and* names of the parent folders indicate the work process in which the series and files were created and the context in which the files and electronic records must be interpretable in the future.

The explicit registration of this contextual information is not only an archival necessity, but it is also a precautionary measure against possible disasters. Loss, for example, as a consequence of transformations is always possible. The folder structure is completely external in relation to the archived records, because they are only preserved at the level of the file system. Except for the filed e-mails (with the embedded metadata), the electronic records themselves do not contain references to the folder structure. Since the electronic records can only retain their function as a record by means of the folder structure, one must in one way or another provide for the registration of the folder structure so it can be reconstructed if necessary. At the latest, this contextual information must be registered at the time when series and files with archival value are moved. There are various possibilities for this.

*XML dossierlists* A first possibility for archiving the contextual information presented by the folder structure and the location of the records within the folder structure in an explicit way, is the creation of a metadata file, called dossierlists. These dossierlists are composed in XML. In this XML document, a structured and explicit statement is made as to how the electronic classification system and its contents was constructed. An XML dossierlist provides a hierarchical overview of the series, files and their records. The nesting of the XML elements reflects the structure and the relationship among the various folders and subfolders. An example of an XML dossierlist is available on the DAVID website<sup>40</sup>.

The compilation of such an XML dossierlist occurs completely automatically. A tool developed specific for the city administration of Antwerp is used for this.

*Replication of the folder structure* Another possibility, when electronic series or files are moved, is the replication of the electronic classification structure from the root down to the level of the selected folder. In that way, the branch of the tree structure of which the selected folder is a part is reconstructed at the temporary location where the transfer to the repository is prepared. In this way, names of functions, series and files are communicated. For this operation, an extension to the Windows explorer was programmed (a Shell Com extension). With this integration, a selected folder can be copied or moved, including the selected parent folder names.



**Illustration 11:** Tool for moving / copying selected folders from the classification system, including the parent folder names reflecting the context.

<sup>40</sup> [http://www.edavid.be/davidproject/nl/xml\\_metadata.htm](http://www.edavid.be/davidproject/nl/xml_metadata.htm).

*Case file metadata* Other archived metadata are the file metadata preserved in the electronic classification system. These metadata are saved in a hidden XML file and serve as a basis for the description of files that are preserved in the metadata system of the digital repository.

Options available when folders are moved from the electronic classification system are an automatic up-dating of these file metadata (with their contents) or the generation of metadata for any file for which they have not yet been generated.

### 4.3 Migration to preservation formats

In the electronic classification system, electronic records are saved in their native application file format. These application file formats are seldom suitable as preservation formats. There is therefore the danger of having a readability problem later when the associated application software is no longer available. As a solution for this digital permanence problem, the DAVID preservation strategy is applied<sup>41</sup>. This strategy is based on migration to suitable preservation formats in combination with the preservation of the records in their original application format. By doing so, various migration and/or emulation options remain open in the future.

#### 4.3.1 ARCHIVING FORMATS FOR E-MAILS AND ATTACHMENTS

*Archiving as XML documents* The Antwerp city archives uses XML as preservation file format for e-mails. The selection of XML is justified by the all-round advantages of XML as a preservation format for electronic records in general. XML is internationally accepted as the most suitable preservation format for e-mails<sup>42</sup>. XML also fits perfectly within the general electronic record-keeping strategy of the city archives, which is based on a minimal IT infrastructure in the administration.

*Document model* For the XML preservation of e-mails, the XML Schema is applied that has been developed by Expertisecentrum DAVID<sup>43</sup>.

---

<sup>41</sup> For the DAVID preservation strategy, see: F. BOUDREZ, *B. Preservation strategies*, in: F. BOUDREZ, H. DEKEYSER AND J. DUMORTIER, *Digital archiving: legal and archival issues*, Antwerp-Leuven, 2004. ([http://www.expertisecentrumdavid.be/docs/digitalarchiving\\_manual.pdf](http://www.expertisecentrumdavid.be/docs/digitalarchiving_manual.pdf))

<sup>42</sup> XML is also designated by the NARA and *Testbed Digitale Bewaring* as the most suitable archiving format for e-mails: [http://www.archives.gov/records\\_management/initiatives/email\\_attachments.html](http://www.archives.gov/records_management/initiatives/email_attachments.html); TESTBED DIGITALE BEWARING, *Van digitale vluchtigheid naar digitaal houvast. Bewaren van e-mail*, p. 36.

<sup>43</sup> <http://www.edavid.be/xmlschemas/email.xsd>

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <email:document xmlns:email="Stadsarchief_Antwerpen/email/1.0">
- <email:sender>
  <email:name>Filip Boudrez</email:name>
  <email:emailaddress>Filip.Boudrez@stad.antwerpen.be</email:emailaddress>
</email:sender>
<email:sent>27/01/2006 10:08:38</email:sent>
- <email:recipients>
- <email:to-type>
- <email:to>
  <email:name>Willem Vanneste</email:name>
  <email:emailaddress>Willem.Vanneste@stad.antwerpen.be</email:emailaddress>
</email:to>
</email:to-type>
- <email:cc-type>
- <email:cc>
  <email:name>Roeland Verhaert SA</email:name>
  <email:emailaddress>Roeland.Verhaert@stad.antwerpen.be</email:emailaddress>
</email:cc>
</email:cc-type>
<email:bcc-type />
</email:recipients>
<email:received>27/01/2006 10:08:38</email:received>
<email:subject>Digitaal depot: systeemvereisten</email:subject>
<email:attachments>Systeemvereisten.odt;</email:attachments>
<email:reference>1_0_Toezicht\4_DigitaalArchiveren\4_2_DigitaalDepot</email:reference>
<email:message>Willem, Als bijlage stuur ik je hierbij een overzicht van de systeemvereisten door. We
zullen dit verder bespreken op de eerstvolgende bijeenkomst van de projectgroep. Ter aanvulling stuur
ik eveneens nog een aantal documenten met achtergrondinformatie door. Mvg, Filip Boudrez |
consulent digitaal archief Stad Antwerpen | Stadsarchief Venusstraat 11 | 2000 Antwerpen tel + 32 03
206 94 11 | fax +32 03 206 94 10 filip.boudrez@stad.antwerpen.be
<mailto:filip.boudrez@stad.antwerpen.be> DAVID-website: <http://www.edavid.be/davidproject>
DAVID-handboek: <http://www.edavid.be/davidhandboek> eFloris-website:
<http://stadsarchief.antwerpen.be/> 't Stad is van iedereen.</email:message>
</email:document>
```

**Illustration 11:** An e-mail preserved as XML-document conforming the eDAVID XML Schema. The XML e-mail contains an explicit reference to the context ('email:reference') and to the filed attachments ('email:attachments').

#### Migration to XML

The migration of the e-mail messages saved as .msg files is done completely automatically. A migration tool has been developed for this purpose. It converts all e-mails to XML one by one. The XML representations of the e-mails get the same file name in the same folder as the .msg files. Only the extension is changed (.xml). This cooperated with MS Outlook for the migration process.

When the .msg files are migrated, the embedded transmission and contextual metadata are retrieved and mapped to the corresponding XML elements. This applies for the e-mail address of the sender, the name and the e-mail address of the authorised delegate, the file names of the filed attachments, the classification or registering reference and the date and time of sending and receipt.

#### Quality control

Ideally, the output of the migration process should be subjected to several quality controls. A systematic and completely automated validation of the XML documents based on the eDAVID XML Schema for e-mails, checks whether the document model was applied correctly. In addition, it is also advisable to have several random manual tests.

### 4.3.2 ATTACHMENTS AND OTHER ELECTRONIC DOCUMENTS

#### Selecting a preservation format

The e-mail attachments and the other electronic documents in the folder structure are not archived as XML documents by definition. The nature of these electronic documents can be diverse. For each type of electronic record a suitable preservation format is used. In this way one also has an immediate solution for electronic records that are not sent as e-mail attachments. It is preferable that the

preservation formats are official standards and not depending to a manufacturer or an application. Important criteria are independence from the software application used to create the documents, and publication of the specifications of the computer file format. The use of compression should be avoided as much as possible. It is best for electronic records to be preserved in a suitable preservation format from the moment of their creation. This is not always possible, however, so some migrations will always be needed. The standards that the Antwerp city archives uses for this are established in *Digital ArchiVing: guldeline & aDvice, no. 4: Standards for file formats*<sup>44</sup>. The Antwerp city archives selected the following archiving formats from this guideline:

Text documents	
MS Word	ODT and TIFF
MS Excel	XML and TIFF, ODS
MS Access	XML and TIFF
Illustrations	
Raster	TIFF (uncompressed)
Vector	CGM
Audio	WAV (uncompressed PCM)
Video	AAF or MXF
CAD	DXF

*Migration to preservation formats* The migration of the electronic records with archival value occurs completely automatically, as with e-mails. To accomplish this, the migration tool for e-mails has been expanded with additional modules so other document types can also be converted.

#### 4.4 Encapsulation in AIP's

Before the records are ingested in the digital repository of the Antwerp city archives, they are first transformed into Archive Information Packages (AIP's). AIP's are the information packages that are managed in the digital preservation system within the OAIS reference model. The Antwerp city archives has adopted the AIP implementation method of eDAVID<sup>45</sup>.

*AIP: metadata, .msg and .xml* In the case of e-mails, this storage method means that the metadata, the message file and the e-mail that is migrated to XML are encapsulated in one AIP. An important metadata element included in the AIP is the location of the electronic

<sup>44</sup> This guideline is an application of DAVID guidelines & aDvice no. 4 (<http://www.expertisecentrumdavid.be/davidproject/teksten/guideline4.pdf>)

<sup>45</sup> Based on the OAIS reference model and on the encapsulation technique, eDAVID developed a storage method in which the essential metadata and the various representations of one record are packed in one AIP container. This container forms one physical entity so the various components of the electronic record are inextricably transferred in time. When essential metadata is present, the digital object immediately has the status of record. These metadata accompany the representations of an electronic record at all times. XML is used here as the encapsulation format. For more information about this storage method: F. BOUDREZ, *Digital containers for shipment into the future*, Antwerp, 2005 (<http://www.edavid.be>).

record in the classification system and the name of the series or the file of which it is a part. By encapsulating these data, the physical folder structure becomes unnecessary, and it is sufficient to maintain one large collection of AIP's.

*Composition of AIP's* The creation of AIP's is also a completely automated process. Depending on the distribution of the responsibilities, this operation can be carried out at the same time as the migration, or one can postpone the encapsulation until a later time. Depending on this choice, the encapsulation can be done by the creating agency or by the archival service. Encapsulation in AIP containers is an optional functionality of the migration tool developed by the Antwerp city archives.

## 4.5 Retrieval and dissemination

*Retrieval: a legal obligation* The last step in the preservation and archiving procedure is making the electronic records retrievable and accessible. Making records public and accessible is a legal obligation prescribed by the freedom of information acts<sup>46</sup>. Actually, this obligation applies both for records in the custody of the creating agency and records that have been moved to the digital repository.

*Options* For the retrieval of electronic records, various options or combinations of options are possible:

- browsing the (virtual) folder structure
- structured searches in the contextual metadata, possibly in combination with:
- full-text searches

The selection of one certain option or even a combination of options depends mainly on which aggregation level the electronic records must be retrievable. Retrieval at case file or subject level is clearly the primary retrieval level. The archivist can accomplish this in various ways: on the basis of XML dossierlists, transferlists and/or on the basis of the case file metadata in which the content of a folder is listed. This can be combined with the encapsulated metadata in the AIP's. On the basis of these contextual metadata, a virtual folder structure can be reconstructed on ingestion in the repository.

*Topic Maps* In the future, the Antwerp city archives will compile an inventory in the form of an XML Topic Map<sup>47</sup> for the retrieval of electronic case files and records, so users can also find electronic documents in some way other than by means of the folder structure. A Topic Map has the advantage that users can retrieve electronic documents using all kinds of associations. The XML dossierlists or transferlists can serve as a basis for the Topic Map(s). Descriptive metadata can supplement these XML dossierlists so dossiers or folders can also be found on the basis of their archival description.

*Searching records* Structured and/or full-text searches in the transmission metadata and in the content of the e-mails can be used for closer access. Once the appropriate series or case file has been found, one can start searching in the records themselves on the basis

<sup>46</sup> See also: *Omzendbrief betreffende het inzage- en afschrijftrecht van de leden van de gemeenteraden, de politieraden, de provincieraden en de raden voor maatschappelijk welzijn met betrekking tot e-mailberichten en geïnformatiseerde stukken*, 28 June 2002. (BS: 19/07/2002).

<sup>47</sup> For more background information about XTM (XML Topic Maps), see: F. BOUDREZ, *XML Topic Maps voor digitale archivering*, Antwerp, 2002 ([http://www.edavid.be/davidproject/teksten/DAVIDbijdragen/XML\\_Topicmaps.pdf](http://www.edavid.be/davidproject/teksten/DAVIDbijdragen/XML_Topicmaps.pdf)).

of certain search criteria (for example, name of sender, date of sending, subject line, etc.). This can be done with a simple search programme that searches through the XML-stored e-mails in a selected folder. Primary retrieval of e-mails on the basis of full-text searches is consciously avoided. Since full-text searches are not always accurate, they result in much noise. Furthermore, for such a retrieval, the development of a central index and the indexing of all archived e-mails is necessary.

## 5. CONCLUSION

---

The careful preservation and archiving of e-mails and their attachments by organisations is not an isolated archiving problem. Preferably, e-mail archiving should be incorporated into a general records management and archiving strategy. If there is no general archiving strategy for electronic office documents, e-mail archiving provides a good opportunity to start developing one.

The proposed archiving solution for electronic office documents in general, and for e-mails and attachments in particular, is closely related to the way administrations preserve paper documents and dossiers. Also in the electronic environment, administrative employees are expected to perform actions such as registration and file creation. These are familiar operations from the paper world that are now carried out in an electronic context.

For judicial and archival reasons, archiving e-mail with the intervention of the end user is the most obvious solution. From a judicial viewpoint, this is the safest solution if one wants to avoid violating the privacy of the sender or the addressee. The intervention of the end user is also required for the selection of the e-mails and attachments with record status, for situating them in a certain business process and for dossier creation.

Thus, the creation of a high-quality archive is not a completely automated process. In the archiving procedure, the filing of e-mails and the creation of case files is a success factor. As in the paper world, both activities require the necessary care, systematics and procedures. The advantage of an electronic environment is that these procedures can be supported better. In this regard it is extremely important to supply filing instruments that are as user-friendly as possible, to incorporate filing mechanisms, and to provide training and instruction. Developing an archiving procedure and integrating records management functionalities within the existing IT environment can help to stimulate this. Only then may one have a reasonable expectation that e-mails and attachments will actually be filed. The filing and archiving procedure, by the way, is not a goal in itself, but benefits operational management and makes accountability possible. It also reduces stress for the administrative employees<sup>48</sup>.

Since the archivist is the architect of the archiving system, he is expected to provide the necessary support.

---

<sup>48</sup> Various stress surveys indicate that a lack of order, and chaotic records management are responsible for stress on the work floor. Long searches for documents lead to annoyance and extra work (for this, see the various stress surveys, the results of which were distributed in the fall of 2005, for example: *Administratieve chaos veroorzaakt stress*, in: *Office Rendement*, 2-9 January 2006).

---

## 6. APPENDICES

---

### 6.1 Tools

For the practical implementation, the Antwerp city archive developed the following tools:

- FilingToolbox:
  - plug-ins for MS Outlook for the capture and registration of metadata and the filing of e-mails and attachments:
    - for individual e-mails
    - for a selection of e-mails
    - for the entire content of an Outlook folder
  - metadata extension of Windows explorer for:
    - the registration of metadata at the series and/or file level
    - the replication of the filing structure when folders are moved for archiving
  - CopyPath: for copying a complete path to folders and/or computer files in Windows explorer
- ArchivalToolbox:
  - migration and encapsulation tool:
    - migration of e-mail (to XML) and word processing files (to ODT and/or TIFF)
    - encapsulation of e-mails, word processing files, images and audio in AIP's
    - automatic updating or generation of series/files metadata.
  - tool for reading AIP's and unpacking representations of electronic records.

### 6.2 Alternative implementations

Building on the DAVID model solution for archiving electronic documents, the Antwerp city archives developed an archiving procedure for e-mails and attachments for the administration of the city of Antwerp. This procedure involves several choices that are inspired by:

- the technological infrastructure: MS Exchange/Outlook as e-mail environment, limited presence of records management applications
- the long-term electronic preservation strategy: the DAVID preservation strategy that combines the preservation of the records in their original application format along with migration to one or more preservation formats
- the storage method in the digital repository: the eDAVID implementation method of OAIS-compliant AIP's
- the vision with regard to metadata

These basic starting points will no doubt differ in one or more aspects from those of other organisations that want to develop an archiving strategy for electronic documents and e-mail. Other creators will be working with different e-mail software or will select a different electronic preservation strategy. In function of their own points of departure they will apply different options or methods. In the following table, various alternatives for (parts of) the strategy of the city of Antwerp are listed. When relevant, possible risks or disadvantages of the alternatives are stated.



	<i>CITY OF ANTWERP</i>	<i>ALTERNATIVES</i>	<i>RISKS/DISADVANTAGES OF THE ALTERNATIVE</i>
<b>ELECTRONIC CLASSIFICATION SYSTEM</b>			
hosted by	shared server disk	e-mail system, DMS/RMA	e-mail system: sharing information with colleagues is difficult, if not impossible + fragmentation of information
location	shared server disk	e-mail system, DMS/RMA	
<b>REGISTERING METADATA ABOUT SERIES AND FILES</b>			
by means of	customisation of Windows explorer	metadata tool, DMS/RMA	
relationship between paper and electronic files	metadata element in XML metadata document	shortcut, flat text file, ... place in folder	
<b>REGISTERING E-MAIL TRANSMISSION METADATA</b>			
when	at the time of filing	at the time of migration	accuracy, availability of e-mail system
storage place	embedding in the filed e-mail	embedding in the preservation format or central database	relationship between the database record and the record may never be lost
<b>REGISTRATION OF E-MAIL CONTEXTUAL METADATA</b>			
when	at the time of filing	at the time of migration	accuracy, availability of e-mail system
storage place	embedding in the filed e-mail	subject or message field of the e-mail, embedding in the preservation format or central database	in the subject line: original subject indication is lost or changed in the message field: original layout of e-mail message is lost electronic signature is unusable relationship between the database record and the record may never be lost
<b>FILING</b>			
by means of:	customisation of MS Outlook	use default "Save as..." functionality, drag to the corresponding folder, provide additional functionalities for other e-mail clients	with "Save as..." functionality: wrong selection of export format is a risk, essential metadata are not explicitly registered, attachments are part of filed e-mail development of an extension/add-on to Mozilla Thunderbird
by:	end user	the records manager	original transmission metadata is lost when an e-mail is forwarded
'filed' status	e-mail has 'filed' as a category designation	marking or labelling of filed e-mails	Thunderbird: create a 'filed' label and assign it to filed e-mails
filing format:	.msg	.txt, .html, .eml	.txt and .html: essential

CITY OF ANTWERP		ALTERNATIVES	RISKS/DISADVANTAGES OF THE ALTERNATIVE
		(Mozilla Thunderbird, Outlook Express, etc.) or immediately in a suitable preservation format	transmission metadata and attachments are lacking immediately in the preservation format: using the e-mail client to consult, answer or forward e-mails will no longer be possible
SEPARATION OF E-MAILS AND ATTACHMENTS			
when	at the time of filing	at the time of migration	looking up and reusing attachments in the active filing system is labor-intensive
how	automatically	manually (when filing) automatically (by migration tool)	manually: more labour-intensive, great chance of errors on migration: a more laborious operation
indicating the relationship	embedded metadata and shortcuts	adapting the file name of attachments	more labour-intensive and chance of errors
ELECTRONIC PRESERVATION STRATEGY			
preservation of:	e-mail in application and preservation format	e-mail in the preservation format	emulation of e-mail in application format is no longer possible migration of e-mail with application format as a source file is no longer possible
preservation format	XML conforming the eDAVID document model for e-mails	PDF/A, HTML, plain text	PDF/A: is PDF/A completely free of patent rights? is PDF/A as simple as XML? HTML/plain text: internal structure is not explicitly established
storage method	encapsulation in AIP's	as separate digital objects	relationship among the objects, and between the record and its metadata may never be lost

### 6.3 Roles and responsibilities

Archiving e-mail is not a matter for which the archivist or the archival department alone is responsible. For the implementation of an archiving policy the following actors are involved: the management, the archivist, the system manager, the LAN manager, the records manager, and the end user. Effective e-mail archiving is only possible when all the involved parties actively participate in the archiving strategy.

#### 6.3.1 MANAGEMENT

- establishes the formal archiving policy of the organisation, including:
- the electronic preservation strategy for long-term preservation

- establishing the roles and responsibilities within the organisation
- provides the necessary time and resources for working out and implementing the archiving policy

### 6.3.2 ARCHIVIST

- designs the general archiving policy for the organisation
- develops an archiving strategy for e-mails within the general archiving policy
- which e-mails are subject to the freedom of information act?
- which e-mails are records: compile a general records schedule, supply selection criteria
- how are e-mails, attachments and electronic documents archived in general?
- how are the context of the electronic records and the mutual relationships archived?
- what happens to the mailboxes of users who leave the organisation?
- identifies the essential metadata
- establishes the filing/export format for e-mail
- establishes the preservation formats for electronic records
- provides assistance or advice when the classification schema is being designed
- takes care of the necessary motivation, training and instruction for e-mail preservation
- makes provisions for retrieval from the digital repository

### 6.3.3 SYSTEM AND MAIL-SERVER ADMINISTRATOR

- sets the security at the e-mail server level (retrieval of the e-mail address of the sender or his authorised representative)
- sets group policy: deletes \*.lnk files from the designated computer file types

### 6.3.4 LAN MANAGER

- installs the Outlook customisations on the client computers
- implements the folder structure
- monitors the quality of the folder structure
- migrates electronic documents to preservation formats
- composes the XML dossierlist of transferlist
- provides technical support for transfer to the digital repository

### 6.3.5 THE RECORDS MANAGER

- designs the classification schema
- establishes the reading and modification rights in the classification schema
- monitors the quality of the classification schema
- registers metadata on series/folders level
- selects the folders with archival value: application of the records schedule

### 6.3.6 E-MAIL USER

- creates archivable e-mails
- registers the contextual metadata for e-mail records

- filing and case file creation: e.g. exports e-mails and attachments with record status

## 7. ABBREVIATIONS

---

BCC	blind carbon copy
CC	carbon copy
COM addin	software extension that is built into an existing software package and that adds one or more new functionalities to it; plug-in based on COM technology
CSS	Cascading Stylesheets
DTD	Document Type Definition
EML	Computer file format for e-mail
ECHR	European Convention on Human Rights
ISO	International Organisation for Standardisation
IT	Information Technology
HTML	HyperText Markup Language
MD5	Message digest algorithm no. 5 (rfc 1321)
MSG	MS Outlook message file
ODT	OpenDocument Format
OFT	MS Outlook template file
PDF	Portable Document Format
PDF/A	Portable Document Format for Archiving
TXT	Flat file
VBA	Visual Basic for Applications
XML	eXtensible Markup Language
XSL	eXtensible Stylesheet Language

## 8. LITERATURE

---

- F. BOUDREZ, *<XML/> and electronic record-keeping*, Antwerp, 2002. ([http://www.expertisecentrumdavid.be/davidproject/teksten/XML\\_erecordkeeping.pdf](http://www.expertisecentrumdavid.be/davidproject/teksten/XML_erecordkeeping.pdf))
- F. BOUDREZ, *Standaarden voor digitale archiefdocumenten*, Antwerp, 2003. (Dutch only) ([http://www.expertisecentrumdavid.be/docs/eDAVID\\_standaarden.pdf](http://www.expertisecentrumdavid.be/docs/eDAVID_standaarden.pdf))
- F. BOUDREZ, *XML Topic Maps for electronic record-keeping*, Antwerp, 2002. ([http://www.expertisecentrumdavid.be/davidproject/teksten/DAVIDbijdragen/XML\\_Topic\\_Maps\\_eng.pdf](http://www.expertisecentrumdavid.be/davidproject/teksten/DAVIDbijdragen/XML_Topic_Maps_eng.pdf))
- F. BOUDREZ, *Digital containers for shipment into the future*, Antwerp, 2005. ([http://www.expertisecentrumdavid.be/docs/digital\\_containers.pdf](http://www.expertisecentrumdavid.be/docs/digital_containers.pdf))
- F. BOUDREZ, H. DEKEYSER and S. VAN DEN EYNDE, *Archiving e-mail*, Antwerp-Leuven, 2003 (Version 2.0). (<http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Report4.pdf>)
- L. DURANTI, *The archival bond*, in: *Archives and museum informatics*, 1997, no's 3-4, p. 213-218.
- *Handleiding archivering elektronische post*, Amsterdam, 2000.

- P. HORSMAN, *Archiveren van elektronische post. Methoden, meningen en alternatieven*, Amsterdam, 1999.
- G. KLYNE, *An XML format for mail and other messages*, 2003.
- TC 46/SC 11, *ISO 15489 Information and documentation -- Records management -- Part 1: General*, 2001.
- TC 46/SC 11, *ISO 15489 Information and documentation -- Records management -- Part 2: Guidelines*, 2002.
- *Managing electronic mail. Guidelines for Kansas Government Agencies*, Kansas, 2001.
- NATIONAL ARCHIVES OF AUSTRALIA, *Managing electronic messages as e-mails. Guidelines*. ([http://www.naa.gov.au/recordkeeping/er/elec\\_messages/](http://www.naa.gov.au/recordkeeping/er/elec_messages/))
- NATIONAL ARCHIVES OF AUSTRALIA, *Managing electronic messages as e-mails. Policy*. ([http://www.naa.gov.au/recordkeeping/er/elec\\_messages/](http://www.naa.gov.au/recordkeeping/er/elec_messages/))
- Model requirements for the management of electronic records, *Moreq specification*, 2002 (<http://europa.eu.int/idabc/and/document/2631/5585>).
- MOORE, R., et al, *Collection-Based Persistent Digital Archives -- Part 1*, in: *D-LIB Magazine*, March 2000. (<http://www.dlib.org>)
- MOORE, R., et al, *Collection-Based Persistent Digital Archives -- Part 2*, in: *D-LIB Magazine*, April 2000. (<http://www.dlib.org>)
- TESTBED DIGITALE BEWARING, *Van digitale vluchtigheid naar digitaal houvast. Bewaren van e-mail*, The Hague, 2003 (<http://www.digitaleduurzaamheid.nl>)
- G.J. VAN BUSSEL, P.J. HORSMAN, H. WAALWIJK, *Softwarespecificaties voor Records Management Applicaties voor de Nederlandse Overheid, ReMaNo 2004*, Amsterdam, 2004.