DOK ■ magazin ■

Technologies, strategies and services for the digital document

Translation of online articles from December 13, 2011 www.dokmagazin.de

The Long-Term Retention of Emails

Dr. Hans Bärfuss, PDF Tools AG

All business leaders today know that a large portion of their organization's value lies in its documents. Documents also form the basis for demonstrating proof in legal disputes. The flood of emails in businesses is now requiring archivists to confront the special demands of retaining digital documents. This challenge is too often left up to the suppliers of archiving systems alone.

In addition, most manufacturers of archiving systems will suggest that creating audit-proof documents can be achieved automatically. Before this can be accomplished though, a whole line of measures has to be taken. For example the archiving format, as well as organizational measures like consolidating emails together with other documents into dossiers, are pivotal. Also, the lifecycle of a retained email is generally considerably longer than the life expectancy of the electronic archiving system. Emails may have to be traceable, faultlessly reproducible and legible after ten, twenty or more years depending on statutory provisions. The long-term retention of emails remains therefore a difficult challenge.

Legal Requirements

Emails are – depending on their content – business records in digital format. The retention provisions are regulated through legislation; in many countries there also exist additional directives for digital dossiers. As a further measure, records management is standardized in ISO 15489 as international "best practice". There are numerous collective requirements, independent of specific legislation in a respective country, that have to be taken into consideration, especially for companies that are internationally active:

- Integrity: Business records must be created and retained in such a manner that they cannot be altered without the alteration being recorded and recognizable.
- Documented processes: The organization, responsibilities, activities and processes that regulate the management and retention of business records, as well as the infrastructure, have to be documented in work instructions.
- Due diligence: Business records must be accurately and orderly retained and protected from damaging influences.
- Availability: It must be possible to view and verify business records within an acceptable time limit. Within this viewing capability there exists also the requirement that it must be possible, upon request of an authorized person, to provide the business records in a format that is legible without aids, i.e. printed on paper. Archived data must also be retrievable within a specified time frame.
- Archiving: Information must be systematically inventoried and protected from unauthorized access, whereby access is to be registered and these registry documents themselves have to be retained in the same manner as the data medium.

- Acceptable data medium: Unalterable information carriers like paper, image mediums (e.g. microfilm) and unchangeable data mediums (WORM). Also acceptable are alterable mediums, provided that the respective technical procedures like digital signature process are used that guarantee the integrity of the information, the date that the information was saved is verifiable (time stamp) and the procedures are documented (recorded in a protocol).
- Verification and data migration: The integrity and legibility of information mediums must be verified on a regular basis. A data migration to another format or information medium is possible, provided the completeness and correctness, accessibility and legibility are still guaranteed, and that a record of the data transfer is made and retained.

The Functionality of an Archiving System

Archive systems save files (objects), relationships between objects (relations) and descriptions of the files (metadata). As opposed to operational DMS systems, archive systems should support retention periods, especially the managed deletion of files, or better said of dossiers, once the retention period has expired. They should additionally guarantee protection from unauthorized access and mutation. Even more: the system should support access supervision through means of a register. Finally, the archive system should ensure safety of the data, for example through a redundant storage of the data and regular copying of the data to a new data medium.

The Choice of Data Formats is Essential

Most manufacturers of archiving systems will claim that the file system used for archiving is irrelevant. Even more: they allege that the requirement of "audit-proof" documentation is fulfilled solely upon the fact that their archiving system is being used. This assumption is not fully correct. The archiving format plays a much larger role than the archive manufacturer would like to admit. The customer should be aware that the life cycle of a document is generally considerably longer than that of the archiving system. As well, in the customer's best interest, the role of the document and that of the archiving system should be clearly separated. For example, proof of a document's integrity, as well as the traceability of changes, belong to the characteristics of the document format.

Based on this awareness, archivists have selected PDF/A as their preferred file format. PDF/A was developed specifically for the long-term retention of digital documents and is the format of choice for most archives in Europe (and predominantly in Germany and Switzerland). The archiving system should ensure that only documents which adhere to this above mentioned format are stored. This can be ensured through use of a so-called "Gate Keeper" in the form of suitable validation software (validator).

Electronic Signature Identifies Changes to a Document

Electronic signatures fulfill four essential functions:

1. They replace handwritten signatures: Within the scope of formal legal requirements with respect to the validity of documents, the electronic signature meets all the prerequisites of, and in fact is equivalent to, an actual handwritten signature.
2. Integrity protection: Electronic signatures have the same effect as a "wax seal" for digital documents, since changes or manipulations to a digital document are both visibly recognizable and can be traced and reproduced. The trustworthiness of a digital document within a business process is thereby increased.
3. Authenticity: An electronic signature on a document guarantees that the person (natural or legal) who signed the document can be identified.
4. Authorization: Rights and privileges can be configured and managed in an electronic certificate, and thereby assigned to a person.

It is important to understand that neither the electronic signature nor a specific document format (whether PDF or TIFF) can prevent a digital document from being either optically or quantitatively changed via technical means. However, the digital signature can ensure that the change is recognizable and reproducible at all times.

The Technical Solution

From a technical point of view, emails display several unique characteristics. They contain attachments in all imaginable file formats, like office documents, CAD drawings and many more. The files are often contained in structured containers (ZIP and RAR), which themselves can be repackaged countless times over. And more: emails often contain several varia-

tions of the same information, for example as pure text as well as HTML files, and can therefore be ambiguous. In addition, emails contain an abundance of information that is not immediately visible, for example the name of the sending and receiving email servers, date and time of reception and further information about the file.
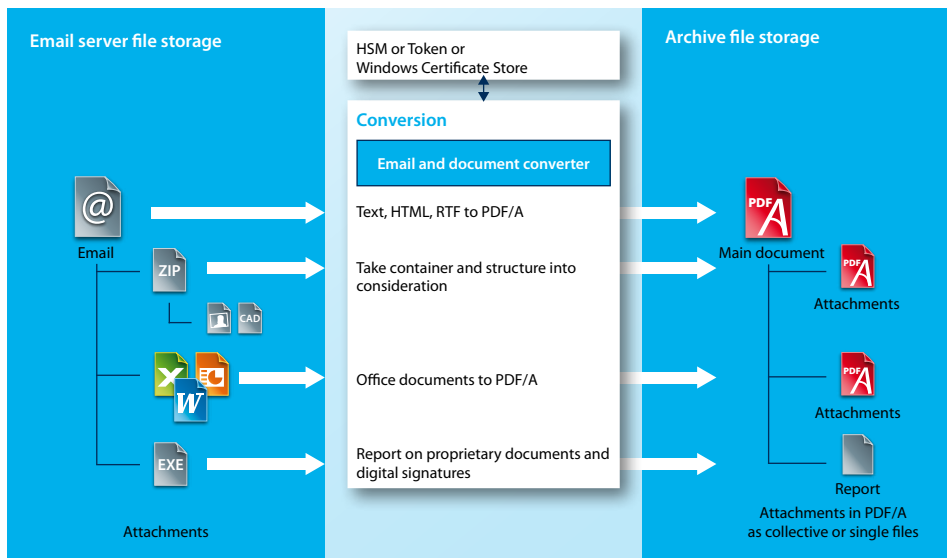

Figure: Converting emails with attachments for long-term archiving

Emails cannot be saved in an archive under these conditions. The hardware devices available on the market (archive appliances) that write emails from a server directly into an "archive" can at best be classified as simple backup devices. Due to the long-term legibility and explicit reproducibility requirements, emails and their attachments have to be converted to PDF/A format, annotated with metadata, and integrated with other digital documents into dossiers and records. When converting to PDF/A, it is vital to ensure that the reproduction of the content in the PDF/A document is carried out in the same manner as the user sees it in the email. This is, due to the above mentioned ambiguous information formats and the multitude of email and office applications and versions in use, not something that can be taken for granted. But it is technically resolvable.

Even though emails have their unique technical characteristics, they can and should be treated like all other documents. Company-wide conversion solutions with product designations like "Document Converter Service" or "Rendition Service" can be referred to for the implementation. Solutions that are dedicated to email archiving alone are in the contrary not recommended: they generally either inadequately or do not at all support the conversion to PDF/A and the consolidation of digital documents to dossiers.

Conclusion

Practically every organization must give thought to archiving emails, whether now or in the future. The legal, organizational, economic and technical aspects of an implementation project must be adequately highlighted and adapted to match the organization's needs. For this reason it is recommended to reserve sufficient funds for consulting services in the preliminary phases of an email archiving project. This groundwork will provide the decision makers in an organization with the right balance between archiving project expenses and the risk of legal actions, giving the project a suitable guideline for the implementation.

Dr. Hans Bärfuss is the founder and CEO of PDF Tools AG, an internationally successful software development and sales company. He represents the Swiss Standards Association at the ISO and actively helps with the standardization of file formats and digital signatures. He is also a founding member of the PDF/A Competence Center, an organization with the objective of promoting the ISO standards and PDF/A. By today, the organization counts more than 100 members, including major companies such as Adobe, Kofax, Nuance and T-Systems. Dr. Hans Bärfuss is the chairman of the Swiss Chapter, speaker at numerous conferences and seminars and publishes articles on the subject of digital documents.