



ArchSafe

Legally secure long-term archiving
of electronic documents

Contact

Your contact for the ARCHISAFE project at the
Physikalisch-Technische Bundesanstalt Braunschweig:

Dir. and Prof. Dr. Siegfried Hackel,
Contact with regard to authorities within the initiative BundOnline 2005
Tel.: +49-531-592-8400
Fax.: +49-531-592-8406
E-Mail: siegfried.hackel@ptb.de

and

Dipl. Wirt.-Inform. Tobias Schäfer,
Project Manager
Tel.: +49-531-592-2456
Fax.: +49-531-592-692456
E-Mail: tobias.schaefer@ptb.de

Introduction

Today, it is hardly impossible to imagine public administration without electronic documents. Due to the broad-scale introduction of IT-supported processing of files (workflow management), files in paper form and the office as the organisational base unit of classical administration increasingly lose in significance. eGovernment, yesterday still a buzzword, is today, without doubt, the embodiment and motor of the model of a future electronic administration. Thereby, it is clear that highly sophisticated electronical front-office innovations on the basis of the Internet are no use at all if in the background, in the so-called "back office", predominantly still telefax, paper binders and file cards are used. The "back office" is the central production site of the administration, the organisation of the "working state", and its main product is the file. Hence, it will also be the file which will be mostly affected by the introduction of modern administrative structures on electronical basis, due to the fact that it has been chosen to be the means which shall enable and support in future the unimpeded and fast exchange of data and information between the different units of the administration, for the benefit of the citizens and of economy - then, however, electronically.

Its main advantage lies in the fact that, as digitalised and encoded information, it can be read directly by machine and that it can, in the form of simple bits and bytes, be transported within seconds, even across large distances.

Digital information is not only volatile (because it is virtual per se), but it can also be manipulated easily and unnoticed. Those who consider the electronic file as being the essential precondition for further progress in eGovernment will have to ponder already today on how the authenticity, integrity, confidentiality and completeness of digital documents can be ensured in the long run, at least for the periods being legally prescribed for the preservation of files. Like their predecessors in paper form, electronic files, too, have to fulfil an evidencing function. This goes far beyond the mere processing of files. Besides, the electronic file will also be subjected to the central "Requirement for the Keeping of Files in Public Administration", which says that it must at all times be possible to gather from the files the complete and true status of an administrative process (case).

The introduction of an electronic document infrastructure remains incomplete if there is no adequate electronic archive. Increasingly, electronic documents are signed. The long-term and legally secure archiving and preservation of electronic information can be ensured with an adequate electronic archive.

The aim of the ARCHISAFE project of the Physikalisch-Technische Bundesanstalt is to develop such an archiving system within the scope of the E-Government-Initiative BundOnline 2005.

Aim of the ARCHISAFE project

ARCHISAFE, the "one-for all" service of the Physikalisch-Technische Bundesanstalt (PTB), supports and furthers the introduction of uniform standards all over Germany for the legally secure and revision-safe long-term storage (archiving) of electronic documents. By establishing a standardized format in XML for the exchange of data (data with regard to the content, description data and signature data) and by implementing software reference architecture, ARCHISAFE creates the essential basis for the introduction and use of both - central and decentral - electronic archives including the Federal Archive.

Together with the BSI (Federal Office for Information Security), the Federal Network Agency, the KBSt (Coordinating and Advisory Agency of the Federal Government for Information Technology in Public Administration), the Federal Archive (Bundesarchiv) and staff from more than 20 different federal (Bund) and state (Länder) authorities, the requirements and technical solution possibilities for a long-term legally secure and revision-safe electronic storage of invoice-founding documents - which fulfil the requirements of the Signature Law (long-term secure storage of the documents and long-term usability) still after a long period of time - have been discussed, assessed and published within the scope of the ARCHISAFE project (see: <http://www.archisafe.de>).

In summary, the following applies:

- ▶ The electronic document infrastructure must - at least - support the long-term archiving of electronic documents, which means that access to electronic documents must - with reasonable effort and expense - also be possible after a longer period of time.

- ▶ The electronic document infrastructure must ensure legal security (in particular the authenticity and integrity of the electronic document) durably - at least, however, until the periods legally prescribed for the preservation of files are running out. This means that compliance of both, images and content, with the original documents must be ensured and that in the case of electronically signed documents, the so-called "re-signature" according to Art. 17 of the Signature Ordinance must be possible.

The decisive aims and principles of the ARCHISAFE project encompass:

- ▶ the use of clearly interpretable, long-term stable and published user data formats
- ▶ the use of clearly interpretable, long-term stable and standardized signature data formats
- ▶ taking into consideration the security suitability of cryptographic algorithms and use of electronic signatures with a sufficiently high security level (qualified electronic signature)
- ▶ the archiving of the required verification data in a form suitable for business operations
- ▶ the timely and probative signature renewal
- ▶ the stable availability of technical components
- ▶ the secure transformation of electronically signed documents
- ▶ the ensuring of data protection and data confidence
- ▶ an increased security thanks to redundancy during the storing and renewing of electronically signed documents
- ▶ cost efficiency due to the possibility of using the data again at a later point in time, due to scalability and by the application of standardised and economical techniques and technologies.

The ARCHISAFE Architecture

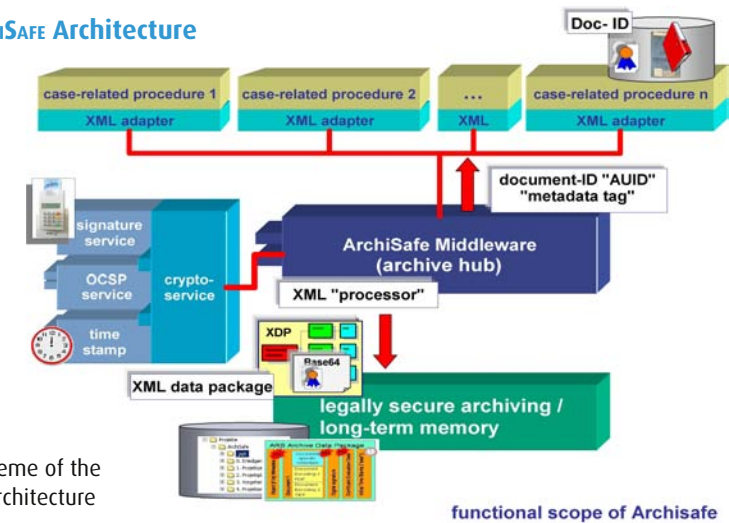


Fig. 1: Scheme of the ARCHISAFE architecture

In technical respect, the ARCHISAFE concept is based on a multi-layer software architecture which is service-oriented and client-compatible.

A so-called "case-related procedure" (e.g. a document-management or file-processing system) serves as a platform and as a leading system for document administration, case creation and interface for long-term storage. The case-related procedure initiates the filing of electronic documents in the electronic archive and administrates the document identifications for the documents stored in the long-term memory. This also implies that the document identifications of stored documents are linked with the document instances and process data stored for the operative processing of a case (see Fig. 1).

The case-related procedure communicates with the long-term memory via a uniform archive interface (archive service) for trans-

mitting the objects to be archived (documents, cases, files) to the long-term memory. The archive service for transmitting the electronic documents from the case-related procedures to the long-term storage system is imaged in a system-independent middleware component (archive hub, see Fig. 1). The archive service checks and processes the archive objects on the basis of standardised character sets and data formats and on the basis of syntactic and semantic agreements for the structures of the data objects to be stored in the long-term memory. Furthermore, the archive service requests, if necessary, cryptographic functions such as signatures, certificate checks and time stamps and processes XML formats on the basis of defined XML-schemes.

The generation of the standardised archive objects and the opening of the communication with the archive service is effected in special service interfaces (service adapters).

The core of the archive service consists of an XML processor with defined interfaces (communication channels) leading to the case-related procedures and to the electronic long-term memory. Furthermore, the archive service shall enable the binding of additional services, such as a signature service (for the generation and/or checking of electronic signatures) and a time stamp.

The cryptographic services sign, when requested by the case-related procedure, the documents which are to be stored in the electronic long-term memory or mark them with a time stamp. In addition, they verify, when requested by the case-related procedure, the signatures and certificates of signed documents and make the verification data available to the archive service which, in turn, then embeds them for a later evidencing function - in a standardised format in the archive object.

In the ARCHISAFE project, the cryptographic services are imaged via the core system of the Federal Virtual Post Office.

The transmission to the electronic long-term memory can, if desired, be combined with a time stamp for the documents to be stored. In addition, a time stamp service is needed for the re-signature of electronically signed documents according to article 17 of the Signature Ordinance (SigV). Here, the ARCHISAFE project builds up on the Archi-Sig-procedure which has been classified as being legally secure and proven (<http://www.archisig.de>). Finally, in the back end, the real long-term memory system is accommodated in which, as a matter of principle, only "original" documents and the associated case meta data are stored. By assigning unambiguous docu-

ment identifications (Document-ID, "Metadata Tags"), the long-term memory system ensures that at any point in time and in economically reasonable time intervals, access is given from the case-related procedure to the stored "originals". This ensures that although the original documents are filed in a legally secure way, the long-term memory system will not be overloaded with case-specific logics. For the realisation of a client-compatible solution, the document can additionally be linked with an unambiguous identification for the respective case-related procedure. In this way, it is possible to ensure, via an authorisation concept in the case-related procedure, that inadmissible access to the archived data is prevented.

In order to ensure access to the long-term memory independent of the case-related procedure, a supplementary search and representation service (on request) can be useful, especially when a client-compatible solution is aspired. Via this service, which keeps the meta data that are stored in the long-term memory redundant by means of a data base, a reconstruction of the cases or files is possible if the leading system falls out. If necessary, the data and documents can be presented (viewer) or exported in such a form that they can be used further. The implementation of such a service must by no means, however, make it possible to undermine the legally prescribed regulations for data protection.

Document formats and data structures in the ARCHISAFE project

Document formats

According to the recommendations of the DOMEA organisation concept version 2.0, only a few formats should be applied for the long-term storage of documents. The co-existence of the most diverse formats in the area of the long-term memory increases the risk that single data types can no longer be reproduced true to the original in the course of the period legally prescribed for the preservation of files and that thus the authenticity of the stored documents goes astray.

ARCHISAFE therefore recommends - on the basis of the DOMEA organisation concept - the following document formats for long-term storage, depending on the format of the data to be archived:

- ▶ TXT (ASCII 7-bit) for simple text information, meta data and master data from special systems,
- ▶ PDF-format (preferably PDF-A) for encoded documents (CI).

This document format can be used independent of the platform and allows - besides storing the graphic information - also the storing of the text information, so that also after the conversion, a full-text search remains possible. Furthermore, the PDF-format has further useful functionalities, such as the embedding of electronic signatures, so that PDF is recommended expressively also by KBSt for the archiving of text documents supplied in CI formats. This is emphasized additionally by the publication of the standard ISO 19005-1 "Document Management - Electronic document file format for long-term preservation - Part 1:

Use of PDF 1.4 (PDF/A-1)"

- ▶ TIFF- and/or PDF-format for documents, which are available in an NCI-format, and
- ▶ XML as markup language for meta data or datasets to be archived.

Meta data structures and interfaces

The "Standards and architectures for E-Government applications" (Vers. 2.0, publications of KBSt, Volume 59, of December 2003) recommend to describe and realize meta data - as well as data interfaces to third systems - as a matter of principle via XML and respective scheme definitions.

Therefore, also for the communication between case-related procedure and archive, ARCHISAFE uses XML as description language for self-contained archive objects which describe themselves via an agreed XML-scheme and thus contain all important information needed for a later access (Figure 2). The description by a valid XML-scheme promises, above all, the following advantages:

- ▶ The archive object can, before being transmitted to the electronic long-term memory, be assessed with regard to its syntactic correctness.
- ▶ Extending the meta data especially for the case-related procedures or for certain public authorities can be achieved with little effort by an extension of the XML-schemes and/or by including additional XML-schemes.

In the simplest case, such an archive object consists, apart from a version number and the indication of the XML-scheme file assigned to it, of a block which contains the data of the content (object block) and, possibly, of one or

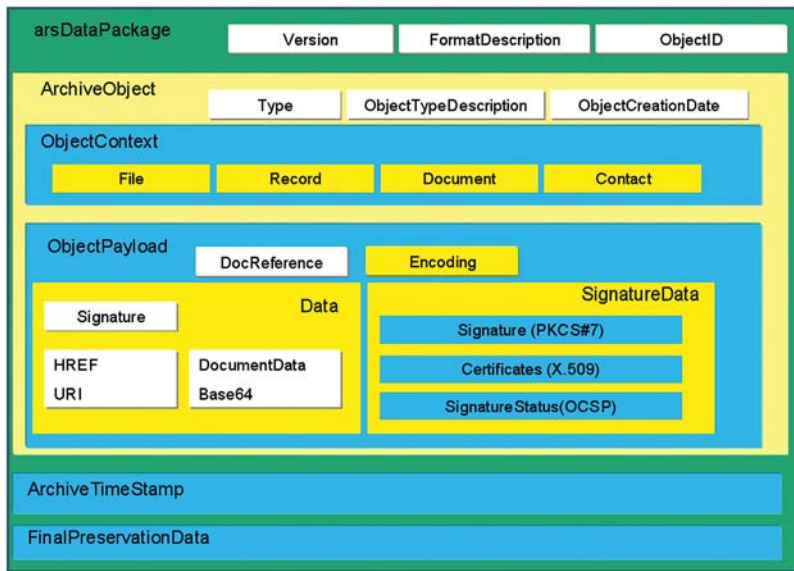


Fig. 2: Scheme of the ArchiSafe meta data (graphical representation))

several signature blocks. The object block itself can have one or several documents which are embedded in XML. Each block contains, as introduction, meta data in which, for example, a document identification (document-ID), a description of the document and of its origin can be stored. As an option, a block for description data is additionally kept available for the transmission to the Federal Archive.

For the document itself, PDF-A is envisaged as standard, which must first be converted into a text format (Base64) before it can be embedded in XML. For storage-intensive binary data it is recommended to reference the binary data as attachment in the XML data flow, not

at last for a better efficiency, in the case of frequent accesses to the long-term memory. In this case, the object block must be provided with a reference to the binary file which is then archived in addition. Furthermore, according to the ArchiSafe-concept, the real content data (document contents) can also be stored in several different document formats inside and outside the XML file. The draft for such an ArchiSafe-compliant XML scheme is specified under the working title ARS (for ARCHISAFE Record-Keeping Strategy) 05.02 "ARS XML data packages and meta data" and has been published in 2005 at www.archisafe.de.

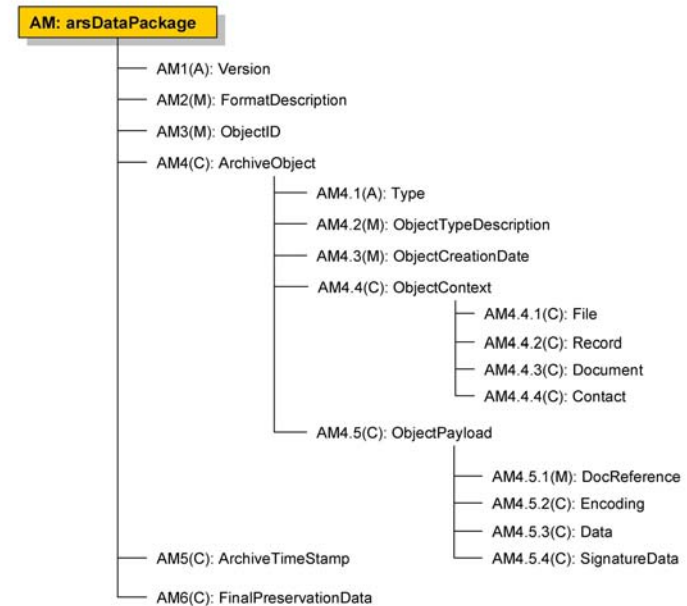


Fig. 3: ARCHISAFE meta data tree (A: attributes, M: meta data, C: container)

In the ARCHISAFE project it is aspired to store a document including the above-described meta data. It must be pointed out here that the relations shown in the DOMEA concept between a file, cases and documents are to be imaged exclusively via the meta data of an "archive object", i.e. of a document. This means that the setting up of a file or a case has to be realized and stored at first outside the ARCHISAFE solution, i.e. in the case-related procedure. By means of the meta data of the "archived" objects, it is possible to re-compile dynamically at all times a list of all the documents that are

stored in the long-term memory and belong to a certain case or file (e.g. via a search). Information on the setting up and development of a file or case can, in addition, be stored via a separate document in the long-term memory (e.g. on the basis of the XDOMEA standard). The projected "single" document/object solution can thus also be used for systems not being compliant with DOMEA, i.e. for any case-related procedures. In addition to that, each subsequent user can define file structures and case structures of his own and enter meta data for this into

ARS, if necessary. Here, the ARCHISAFE project builds up on the experience having been gathered in the federal state of Lower Saxony at the IZN (Centre for Information Technology).

The ARCHISAFE project is developed in three stages. In the first stage, a technical concept and a data processing concept have been developed. On this basis, a pilot has been installed at PTB in the second stage which, although still limited in its functions, is fully operative and makes it possible to gather experience especially with regard to the realisation and embedding into the electronic document infrastructure of PTB and to offer the ARCHISAFE solution as a service also to other public authorities.

Current state of the project Further information

For 2006, the following further developments are projected, among other things:

- ▶ connecting the Federal Virtual Post Office (VPS) with the ARCHISAFE Middleware for obtaining
 - signatures, certificates and information for the checking of signatures,
 - time stamps
- ▶ Conception and implementation of ARCHISAFE's client compatibility
- ▶ Conception and implementation of a secure access channel to the ARCHISAFE-service (manual and automatic call)
- ▶ Completion of the ARCHISAFE specifications
- ▶ Checking of the possibility of certifying the component used within the scope of ARCHISAFE
- ▶ Drawing up of a process documentation according to the principles of due DP-supported accounting systems (GoBS)
- ▶ Development of an "ARCHISAFE" extension module for the DOMEA organisational concept in cooperation with KBSt
- ▶ Setting-up and management of an ARCHISAFE service for external clients

The concepts, specifications, interface definitions and experience reports are published for further use and discussion in an online forum which has been established especially for that purpose and is freely accessible (<http://www.archisafe.de>). Due to the fact that more than 20 federal authorities are bound in - among them the Federal Archive, KBSt, BSI and the Federal Network Agency - a wide field of application opens up for further use in accordance with the "one-for-all"-services promoted by the BundOnline 2005 initiative.

Partners of ARCHISAFE



This flyer was supported by
Micus Management Consulting GmbH, Düsseldorf, Berlin, www.micus.de

