Translation
0495/07

Open

Document Exchange Formats

for the

Federal Administration

Paper for discussion

Berlin and Bonn, February 2007

Contents

1 Electronic documents in the federal administration

Electronic documents are a central tool in the federal administration and are used in two contexts:

- **Editing** documents. Editing includes both the drafting of the original document, the subsequent modification and the passing on of documents to other users and organizations for the purpose of further processing or reproduction.

- **Archiving** documents. Archived documents are no longer produced, but merely reproduced by displaying or printing. Archiving is understood here not just in the technical sense as time-limited storage but also as permanent safe-keeping and availability with a view to legal archiving requirements.

Due to the central role which documents play in the organization and documentation of work by the federal administration, the properties of these documents, or more precisely the document exchange format, are extremely important[1].

The federal administration almost exclusively (currently: approx. 95%) uses Microsoft Office software suites[2]. The related document formats are "closed", i.e. are not fully or not regularly published and are exclusively controlled by the software manufacturer, in this case, Microsoft. This is why the development of independent software[3] for processing Microsoft Office documents is difficult. Moreover, today's widespread use of active contents in documents (e.g. macros) hinders interoperability.

The consequence is that Microsoft Office suites always have to be used for optimum processing of documents, and even always in the same software version if possible. If different software is used for editing, contents and document properties may be lost or accidentally changed, especially with regard to the graphic presentation of the document.

This problem is broken down into different sub-aspects:

- **Poor "tool interoperability" hinders the passing on of documents.** In practical terms, it only makes sense to pass documents on to partners who use the same software. Other individuals, organizations or organization units, however, often use one of the many other software packages.

- **Poor "generation interoperability" hinders archiving.** In order to ensure that the documents of a stock that has grown over the course of years remain legible, a host of software versions must be kept at the ready for accessing these documents. This means that under certain circumstances it may be necessary to maintain operating systems or even hardware that would otherwise be long discarded and which are no longer supported by the respective manufacturers. This goes hand in hand with costs and risks. The subsequent process of

1 The terms "document exchange format" and "document format" are used in this paper as synonyms.

2 Source: IT-InfoBund

3 "Independent software" here means software that uses neither Microsoft Office programming interfaces nor information made accessible by Microsoft within the scope of special contracts only.

conversion into a special long-term, stable document format which is hence necessary today is expensive and there is also a risk that information and functionality may be unintentionally lost.

- **Tool and manufacturer loyalty hinders reliable access for the author to his/her document.** Viewed from a distance, the technical difficulties described mean that the author and the intellectual owner of a document do not have reliable access to their product, instead they are forced to rely on their technical skill and, above all, the development and intellectual property policy of software manufacturers.

If, in addition to manual editing of documents by human authors, automated editing is also considered, another aspect then becomes obvious:

- **Poor tool interoperability makes modern eGovernment more difficult.** If it is difficult to edit a document using alternative office software, then it is also difficult for programs to automatically edit documents. However, in the interest of boosting the efficiency of internal workflows in the public administration and improving service quality for citizens, it must be possible to automatically edit documents. The lack of interoperability between the different office software thus hinders two central aspects of modern eGovernment.

- **Tool and manufacturer loyalty hinders reliable access for citizens to administration information.** Access to a document is not just important for the author himself/herself. Manufacturer-specific formats can also pose difficulties when it comes to communicating with citizens. This holds particularly true with regard to the statutory right to access administration documents in public archives after a protection period of 30 years.

2 Current document formats

The market for document formats which support the functional scope of conventional office program suites is developing at a dynamic pace. The following document formats are attempting to solve the problems referred to above:

Open Document Format (ODF)

- Area of application: Editable format for exchanging documents
- Standardised by: OASIS (Organization for the Advancement of Structured Information
- Implementations: e.g. OpenOffice 2.0; Star Office, IBM-Workplace-Shell, TYPO3,
- Major commercial sponsors: sun microsystems, IBM
- Scope of specifications: Approx. 700 pages
- Cost of specifications: At ISO: 340 CHF, at OASIS: free of charge

Office Open XML (OOXML)

- Area of application: Editable format for exchanging documents
- Standardised by: Ecma International (planned: ISO)
- Implementations: MS Office 2007
- Major commercial sponsor: Microsoft
- Scope of specifications: Approx. 6,000 pages
- Cost of specifications: At Ecma International: free of charge

Portable Document Format Archive (PDF/A)

- Area of application: Non-editable format for document exchange; specially for archiving
- Standardised by: ISO (Nr. 19005-1)
- Implementations: e.g. myPDFconvert (Detec); PDFlib; ghostscript
- Major commercial sponsor: Adobe
- Scope of specifications: Approx. 1,000 pages
- Cost of specifications: At ISO: 114 CHF

3 Properties of open document formats

Open document formats could go a long way towards solving these problems because they comply with the following requirements:

- **Open document formats are independent.** An open document format is standardised and can be implemented by users in their office software and used in any manner free of charge and without any property right restrictions. The user is not dependent on a certain software product.

- **Open document formats are developed in an open process.** The standards are defined by a host of interested parties, e.g. the software manufacturer, significant user groups and independent experts. Access to this process is open to anybody with an interest in the matter. Decision-making processes are transparent.

- **Open document formats are sufficiently documented.** The documentation is available to everybody and the formats can be integrated into the users' own products. A reasonable once-off fee may be charged for documentation.

4 Advantages of open document formats

Open document formats offer a host of advantages which help to solve the problems described earlier:

- **Competition and diversity.** Open document formats make it easier to develop software for interactive and automatic editing and processing, as well as the reproduction of documents. This promotes software diversity, innovation and competition.

- **Interoperability and automation.** At the same time, interoperability is established between the many different software packages, so that it becomes easier to pass on electronic documents, thus enabling, for instance, the automation of eGovernment services.

- **Archive security.** Furthermore, long-term archiving of electronic documents becomes more secure because, if necessary, new software can be produced for document access if the software originally used is no longer available or can no longer be used.

- **Future proof.** Defining standards always calls for a long-term strategy. Documentation with open access and open development processes effectively hinders "digital memory loss".

These benefits are the reasons why the federal administration will use open document formats in the future.

5 Introduction of open document formats in the federal administration

In future, the federal administration is to introduce open document formats throughout the administration. This applies both to documents about to be edited and to archived documents.

In addition to this, the federal administration will launch the following measures:

- **Development of a strategy for the introduction of open document formats.** The federal administration will submit a phase plan for the introduction of open formats. This will focus on the following aspects:

  - Stock-taking and practical examples
  - Evaluation and technical consolidation
  - Pilot operation
  - Mixed operation
  - Migration of legacy data
  - Regular operation and development of prospects

  One challenge for regular operation will be the central handling of technical matters related to interoperability and conformity as well as the openness of formats. Responsibility for this will have to be placed in capable hands.

- **Integration into administration workflows, as well as into standards, rules and recommendations.** In addition to technical issues in conjunction with the introduction of open formats, this development must also consider general administration workflows. The introduction of open document formats, for instance, has implications for invitations to tender and procurement concerning IT systems and products. The DOMEA concept for document management, for instance, and electronic archiving in the public administration and the SAGA paper[4] must still be adapted. After all, this must be considered in all legislation procedures which deal with the electronic exchange of documents.

- **Broad national and international co-ordination.** The federal Government is committed to regular co-ordination with key institutions at municipal, federal-state and EU level in order to exchange experience and to identify and make use of possible synergies.

[4] Standards and Architectures for eGovernment Applications

The Interdepartmental Coordinating Committee for Information Technology (IMKA), the Co-operation Committee for Automatic Data Processing (KoopA ADV) within the scope of co-operation between the federal government, federal states and municipalities, along with the European Commission's IDABC programme, have a central role to play here. Part of this co-ordination involves establishing communication platforms where information can be exchanged.

- **Commitment to the development of open document formats.** The federal administration is committed to the processes of developing and standardising open document formats in order to generally promote open document formats and to especially ensure that the requirements of the public administration are considered. It will support the work of standardisation committees, such as DIN.
  Other tasks include identifying a suitable committee to examine the interoperability and conformity of formats so that maximum interoperability can be guaranteed.

This paper was drafted by a working group comprising representatives from the following ministries and agencies:

- Federal Foreign Office
- Federal Ministry of Economics and Technology
- Federal Ministry of the Interior
- Federal Archives
- The Federal Commissioner for Data Protection and Freedom of Information
- German Federal Office for Information Security