

## ARCHIVING ELECTRONIC RECORDS ACCORDING TO THE OAIS MODEL

### An Organizational Challenge

**Martin Stürzlinger\***

**UDK: 004:930.25**

**Martin Stürzlinger: Arhiviranje elektronskih dokumentov v skladu z modelom OAIS - organizacijski izziv. Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja, Zbornik referatov z dopolnilnega izobraževanja, Maribor 8/2009, str. 405–420.**

*Izvirnik v angleščini, izvleček v slovenščini in angleščini, povzetek v slovenščini.*

Članek podaja pregled najpomembnejših standardov za zaščito elektronskih podatkov - OAIS (odprt arhivski informacijski sistem). Večina opisanih funkcij predstavlja organizacijski izziv, kjer so v ospredju dogovori z ustvarjalci dokumentov, kontrolni sezname zahtev uporabnikov in nadzor. Članek predstavlja nekatere teh izzivov in kot nadaljnjo informacijo opozori na nekatere pomembne projekte in iniciative.

**UDC: 004:930.25**

**Martin Stürzlinger: Archiving Electronic Records According to the OAIS Model - An Organizational Challenge. Technical and Field Related Problems of Traditional and Electronic Archiving. Conference Proceedings, Maribor 8/2009, pp. 405–420.**

*Original in English, abstract in Slovenian and English, summary in Slovenian.*

This article gives an overview of the most important standard for the preservation of electronic data - OAIS (Open Archival Information System). Most of the described functions are an organizational challenge where agreements with records producers, checklists for user requirements and audit trails are in the foreground. Some of those challenges are explained and links to the most relevant projects and initiatives are given for further guidance.

*Ključne besede: OAIS, e-arhiv, standardi, priporočila, Scope.*

In this article I want to give an overview of the OAIS model and show some of the challenges archives are faced with when implementing it. The description of OAIS will be rather short giving references to more elaborate explanations where many important details are explained. For this publication I will focus on certain aspects that I found to be overlooked in some of the electronic records projects that I have found in place.

As most projects also preservation projects usually start with a very limited scope. After the first evaluation period the findings are generalized to the whole system. This approach carries the danger of overlooking many challenges. The initial project usually is simple and workable and only includes one or two producers with limited variety of files to transfer. But when 50 or more producers with a wide variety of records are involved the task is very complex. This article should point out some (and only some) of these challenges.

---

\* Dr. phil. Martin Stürzlinger, MAS, MLS, Scope Partner, Albertgasse 33/17, A-1080 Wien, Österreich.

## THE OAIS (OPEN ARCHIVAL INFORMATION SYSTEM)

Archiving and long-term accessibility of electronic documents is the greatest challenge of our days in the field of archives. Since our information society is built on electronic documents, their preservation and future use is also a critical question of general interest.

The archives community has realized this problem with some delay and has been working on concepts. Since a couple of years the focus has shifted from concepts to finding and implementing solutions. It started with individual projects by archives with enough foresight and resources. Today more and more archives are working on or with more general solutions.

The archives can only control to a certain degree the production and the amount of electronic records produced in their administration. Rarely has the archivist a chance to influence the form these records are produced in.<sup>1</sup> Almost every administration has a different form for its electronic records; very often the understanding of content and form of a record is fundamentally different. Standards are only slowly beginning to change this field.<sup>2</sup>

Just as diverse as the systems on the side of the producers of archival material are the solutions to import and archive on the archives side. While the archives world has traditionally been a field not touched by standards, things have changed in the last 20 years. We got used to the more general standards like ISAD(G) but slowly are moving to more and more precise ones like EAD or PREMIS.

The role of standards is often not appreciated or underestimated. Standards limit the flexibility of a project and will probably not allow for meeting 100% of the organizational needs. Standards are seen as a hindrance when their scope does not exactly match the project goals. But on the other hand standards can also save a lot of work as they stop us from reinventing the wheel. They are mostly drafted by experts who have a wide understanding of the field. Standards should work as guidelines to make sure that no important issue is overlooked. In any case standards are a must if interoperability is part of the project as it is the case for Ingest, archival description and Access.<sup>3</sup>

No general opinion has been found on how to tackle the challenge of archiving electronic records - indeed there are still very few "best practices". Projects and solutions are as diverse as the archives community. Only one standard is generally agreed upon which sets the stage for archiving electronic records. The standard OAIS (Open Archival Information System)<sup>4</sup> is a reference model for a long-term preservation archives. It is however not an implementation guideline.<sup>5</sup> It describes

---

<sup>1</sup> This is at least true for most of the central European countries; it is different in the countries with the tradition of record managers who work as intermediates between the producers and the archives.

<sup>2</sup> I will only point to MoReq2: Model Requirements for the Management of Electronic Records (2008): [www.dlm-network.org/moreq2](http://www.dlm-network.org/moreq2).

<sup>3</sup> Sabine Schrimpf, *Standardisation in the Area of Digital Long-Term Preservation*. Archiving 2008, June 24-27, 2008, Berne, Switzerland.

<sup>4</sup> The standard can be found at: <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

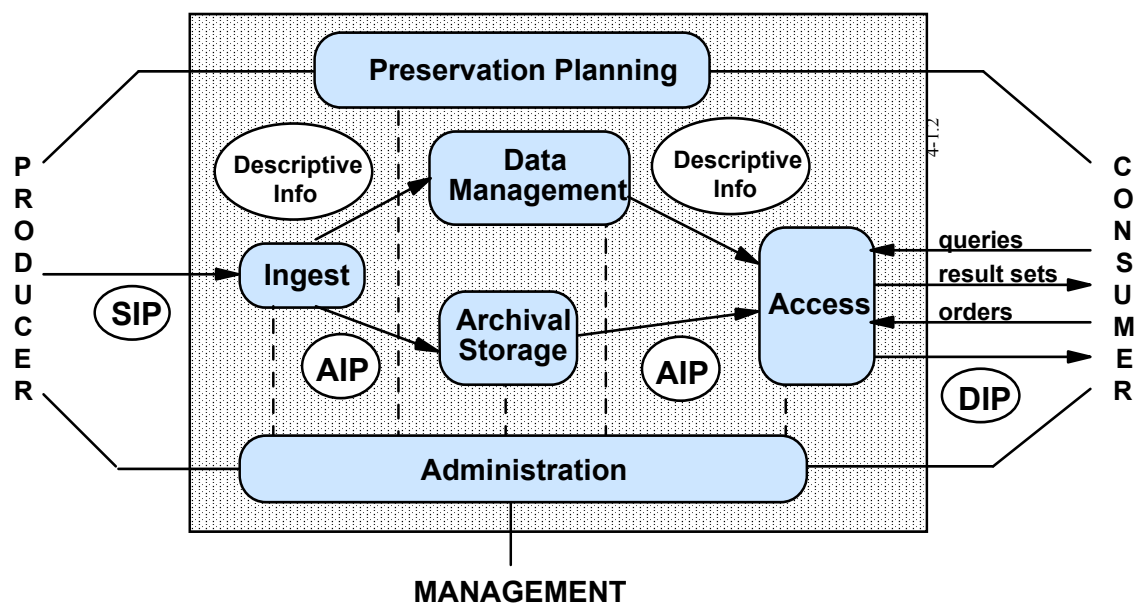
<sup>5</sup> Alexander Egger, *Shortcomings of the Reference Model for an Open Archival Information System (OAIS)*. TCDL Bulletin 2006, Volume 2 Issue 2: <http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html>.

several functions between an OAIS (an archive following the standard OAIS), its management as well as record producers and record consumers.<sup>6</sup>

### SCOPE OF THE STANDARD

When looking at the standard one has to bear in mind that it is not a construction manual for an electronic archives. Not all the functions described need to be solved by one software system but can be solved by several components working together. Some of the functions do not need to be supported by a database at all. They can be organized in paper; like signed agreements between the producer and the archive.

It is important to remember that the challenge of archiving electronic records must first be dealt with on the organizational level and then secondly on the technical side. This is especially true for the Ingest when records are transferred to the archives. The following graphic gives an overview of the model where the most important entities are shown. In this article I will only refer to these parts leaving much of the details aside. The whole OAIS standard is quite complex but still an interesting document. It gives the reader a deeper understanding of the necessary components and how things can work together.



### INFORMATION PACKAGE

The OAIS model describes three different information packages:

1. The Submission Information Package (SIP) is sent from the information producer to the archive, the

<sup>6</sup> Most initiatives and projects for long-term preservation have adopted OAIS as their general reference model, for example the American National Archives (<http://www.archives.gov/era/>), the National Archives in London (<http://www.nationalarchives.gov.uk/preservation/digital.htm>), or the Australian National Archives (<http://pandora.nla.gov.au/>).

2. Archival Information Package (AIP) is the information package actually stored by the archive, and the
3. Dissemination Information Package (DIP) is the information package transferred from the archive to a user in response to an access request.

In the context of preservation the relevant information package is the AIP, since this is the package which is retained over the long-term. Each information package needs four information objects: (1) information about its content - like the documents of a record, (2) a description about the content to make the package searchable and findable, (3) information that is needed for preserving the content, including a unique identifier and the proof of authenticity and (4) information about the whole package that keeps the object together.

## **PRODUCERS**

Producers are all institutions or persons who transfer material of archival value to the archives. Usually a legal basis or a contract regulates the relationship between them. The producer controls the creation process of the records and determines the stage in the lifecycle when the records are transferred to the archives. It is the producers' responsibility to make the export from the original system and create the SIP. The agreement between archives and producers has to make sure that SIP is created according to an agreed interface and transferred in agreed intervals and quantities.

## **INGEST**

The information in electronic records comes from the producers in the form of a SIP, is ingested into the archives, transformed and then stored as an AIP. Most important the information has to be readable and understandable over the long-term and has to remain authentic. To ensure this a number of checks must guarantee the relevant quality criteria.

After the data has been determined complete and is ready for archiving it is sent on to the next steps in the process. First only the descriptive metadata is stored in the Data Management and second the AIP is stored in the Archival Storage.

## **DATA MANAGEMENT**

In the Data Management all descriptive data about the electronic documents are stored. In addition the access rights and protection terms are administered. The data is also the basis for reports, statistics and any query on the data. This function is the part close to the traditional archival finding aid. Advanced archival information systems (AIS) like scopeArchiv can manage the functions in Data Management.

## **ARCHIVAL STORAGE**

The Archival Storage of the OAIS has the safeguarding and preservation of the data at its core. Here the AIPs are stored - they contain all the information; that is the descriptive metadata, the content with all its electronic documents (e.g. the PDFs and TIFFs), the preservation metadata and the information that holds the AIP together. Error checking, data recovery and migration/emulation are handled here.

## ADMINISTRATION

The administration is only in part a job for software but is about organizational issues. This part deals with the agreements for the preparation and ingest of the SIP, the rules for access and the overall standards and policies.

## PRESERVATION PLANNING

The preservation planning monitors the changes in technology and develops the strategy that leads to the migration/emulation of the data.

## ACCESS

The Access provides for all the customers the possibility for searching and retrieving the archived data. According to the OAIS model it is not only an internet portal but also covers the physical reading room where people consult finding aids in paper. In Access the consumers request archival material which is delivered as a DIP in a form suitable for the reader. Rights must be checked if the consumer has the according privileges to view a certain file and billing must be handled. Access should be a function handled by a modern archival information system (AIS).

## DESCRIPTION OF FUNCTIONS, STANDARDS AND SYSTEMS

It is important to see that many of the requirements of long-term preservation of electronic records are organizational needs. Only when those are met then the implementation of technical systems should start. In the following chapters I will point out some of the issues that have to be agreed upon.<sup>7</sup>

Since the number of challenges and difficulties are quite large it is advisable to start with pilot projects that have a limited scope. Invaluable lessons can be learned from actually trying to do things. However for large scale projects at least the requirements for the whole system should be clear. It can prove to be very costly for example to implement an Ingest process without defining the rules and needs for Access. The amount of planning for a full scale system should never be underestimated.

## DEFINITION OF THE INFORMATION PACKAGES

One of the most important concepts in the OAIS standard is the definition of the three different types of information packages; (1) the SIP that is transferred to the archives, (2) the AIP that is stored for long-term preservation and (3) the DIP that is used by the consumer. The information package contains the data about the content itself (for example the TIFF-files) but also the information on how to read and preserve it. As a result the content files must be surrounded with a whole range of metadata. The OAIS standard only gives some indications on what this metadata must fulfill, leaving the details to other initiatives.

---

<sup>7</sup> A very useful starting point on terms and definitions (for German speakers) is: *nestor - handbuch - Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 1.5 (Dezember 2008): <http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf>.*

For some time after OAIS had been published the specification of an information package has been unclear. In the last 5 years quite a number of national and international proposals have emerged on the exact form of such a package. Quite a few projects have implemented these and gained firsthand experience. Those lessons learned must be examined closely to ensure a successful preservation project.

On a national level there exists several definitions for an information package, like in Austria the EDIAKT II standard,<sup>8</sup> in Switzerland the GEVER standard for the SIP,<sup>9</sup> in Germany the very specific DOMEA-Aussonderungskonzept<sup>10</sup> or in France the “Standard d'échange de données pour l'archivage”.<sup>11</sup>

All of these are close to the records management standards in the respective countries and are very suited for the transfer of the data to the archives. When it comes to storing or disseminating the information these standards might not be as adequate. On the international level several initiatives have created other standards that have had great impact on the discussion. These standards fall in two distinct categories, one for the archival metadata and the other for the technical metadata. For the first category METS, EAD, the Dublin Core Metadata Initiative, DIDL and XFDU<sup>12</sup> are examples, for the second there is PREMIS but also LMER.

METS stands for Metadata Encoding and Transmission Standard and was designed to be used as an information package according to OAIS.<sup>13</sup> It is a standard for encoding descriptive, administrative and structural metadata for digital objects. A METS package for a digital object contains information on the metadata associated with the object, the structure of the object and names and locations of files that comprise the object. Other metadata standards such as PREMIS, EAD and MIX are compatible with METS.<sup>14</sup>

METS and PREMIS have emerged as the most important standards in this area. They basically regulate different issues but are closely connected. PREMIS can be embedded into the METS object, a definition for this has already been found.<sup>15</sup> However some projects have elected to use PREMIS as the standard for the whole

<sup>8</sup> *Empfehlung für das XML-Schema zu EDIAKT II*: <http://www.ag.bka.gv.at/index.php/Portal:Ediakt>.

<sup>9</sup> *KOST.07 Archivierung von Unterlagen aus Geschäftsverwaltungssystemen (AUGev) Zusammenfassung: Datenmodellierung und Schnittstellenspezifikation*: <http://www.kost-ceco.ch/cms/download.php?3d502ef689ec5e4ab67b5ed3f6e118a0>.

<sup>10</sup> *DOMEA Konzept - Organisationskonzept 2.0 - Erweiterungsmodul zum Organisationskonzept 2.0 Aussonderung und Archivierung elektronischer Akten - Schriftenreihe der KBSt 66, Oktober 2004*: [http://www.verwaltung-innovativ.de/cln\\_117/nn\\_1007474/SharedDocs/Publikationen/DE/domea\\_konzept\\_aussonderung\\_und\\_archivierung\\_elektronischer\\_akten.html?\\_\\_nn=true](http://www.verwaltung-innovativ.de/cln_117/nn_1007474/SharedDocs/Publikationen/DE/domea_konzept_aussonderung_und_archivierung_elektronischer_akten.html?__nn=true).

<sup>11</sup> *Standard d'échange de données pour l'archivage. Transfert - Communication - Elimination - Restitution (also available in English)*: [https://www.ateliers.modernisation.gouv.fr/ministeres/projets\\_adele/a103\\_archivage\\_elect/public/standard\\_d\\_echange\\_d/](https://www.ateliers.modernisation.gouv.fr/ministeres/projets_adele/a103_archivage_elect/public/standard_d_echange_d/).

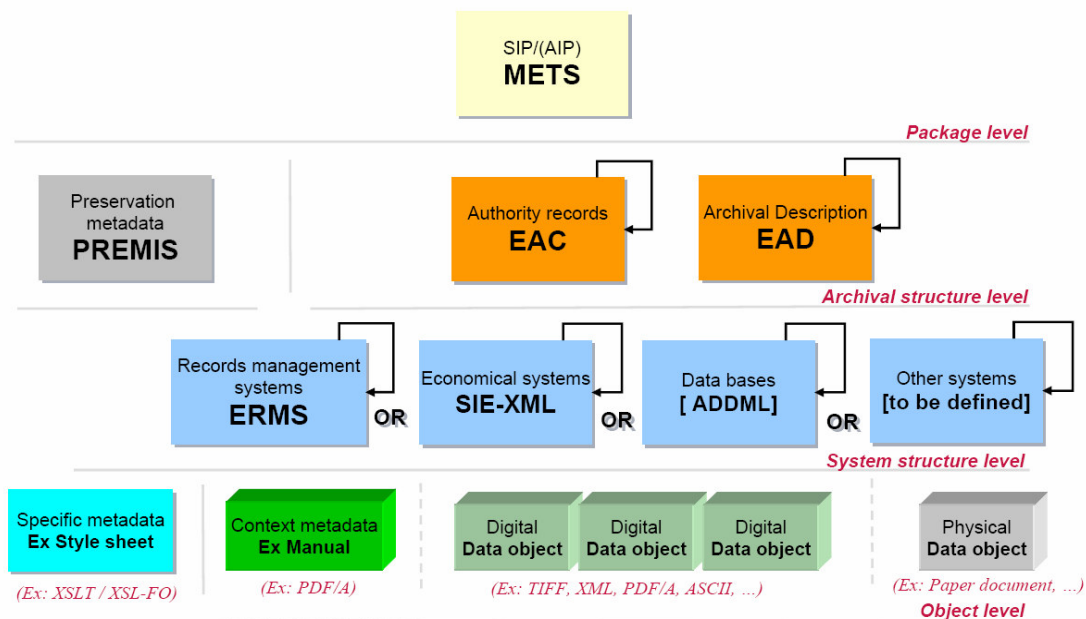
<sup>12</sup> *The XFDU standard is very new: XML FORMATTED DATA UNIT (XFDU). Structure and Construction Rules. CCSDS 661.0-B-1 (September 2008)*: <http://public.ccsds.org/publications/archive/661x0b1.pdf>.  
*Lucas Arnaud, XFDU packaging contribution to an implementation of the OAIS reference model (November 2005)*: <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/043.pdf>.

<sup>13</sup> *The Standard can be found at*: <http://www.loc.gov/standards/mets/>. *A good graphical overview can be found at*: <http://sunsite3.berkeley.edu/mets/diagram/#a3>.

<sup>14</sup> *See also: nestor - handbuch - Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 1.5 (Dezember 2008), Chapter 10: Standards und Standardisierungsbemühungen.*

<sup>15</sup> *Guidelines for using PREMIS with METS for exchange (17. 9. 2008)*: <http://www.loc.gov/standards/premis/premis-mets.html>.

information package. More and more projects elect to use EAD for the archival description.<sup>16</sup> The possibly complexity of an information package is shown in this graphic:<sup>17</sup>



## INGEST

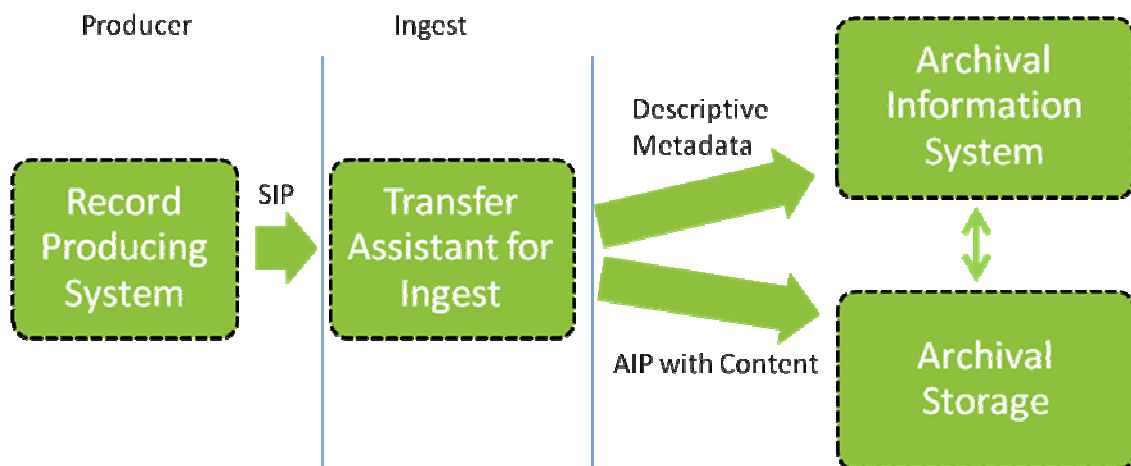
As mentioned the OAIS model provides the general framework for the Ingest process. As a reference model it only gives broad and abstract definitions on the necessary steps and functions to be performed during Ingest. From this basis it is quite difficult to build a working organizational and technical system. To master this challenge several working groups on an international or national level have compiled valuable guidelines giving detailed definitions and descriptions of the necessary steps.<sup>18</sup> The Ingest is the first entity described in the OAIS model and should also be the first part to be solved in a long-term preservation project. Within a data management or recordkeeping system records are produced by an administrative unit - the producer. This data is kept for a specified period of time (usually 5-15 years) until these records are no longer needed in the productive system. Most of the records will be deleted; some will be needed for long-term preservation in an archive.<sup>19</sup>

<sup>16</sup> For example: KOST-Empfehlung zur Archivierung von Unterlagen aus Geschäftsverwaltungssystemen: Prozesse, Datenmodell, Schnittstellen (10. 9. 2008): [www.kost-ceco.ch/cms/download.php?802e473c1bcf4ce83e6e30a93cb78815](http://www.kost-ceco.ch/cms/download.php?802e473c1bcf4ce83e6e30a93cb78815)

<sup>17</sup> Magnus Geber, Developing long time preservation solutions - experiences from the LDB-project and other activities: <http://www.cen.eu/CENORM/BusinessDomains/businessdomains/iss/activity/presldb0706.pdf>.

<sup>18</sup> CCSDS, Producer-Archive Interface Methodology Abstract Standard - PAIMAS (May 2004): <http://public.ccsds.org/publications/archive/651x0b1.pdf>.  
Trade Facilitation and Business Process Working Group: 19 eGovernment, Record Exchange Standard Version: 1.0 Release: 1 Date: 23 June 2008 (Business Requirements Specification BRS, Business domain: Archiving and records management, Business process: Transfer of digital records Document identification) Wege ins Archiv. Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv, Version I. Hg; nestor-Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff (nestor-materialien 10): <http://nbn-resolving.de/urn:nbn:de:0008-2008103009>.

<sup>19</sup> A very interesting initiative is the Archives Ingest Cooperation Wiki where many European national archives work together on describing the ingest process: <http://www.eha.ee/wiki>.



## APPRAISAL

Appraisal must be part of the workflow of records into the OAIS. The majority of the appraisal decisions will be taken before the export from the recordkeeping system. The recordkeeping system must store the information about the archival value of each record in order to facilitate an automated export. In addition to the archival value each record also needs the information about the retention period. Those records with only a short retention period and of no archival value will not be exported but kept in the originating system until their retention period has expired.

All other records with a longer retention period (more than 7-10 years) will have to be exported from the originating system at one point in time. All records of archival value should be transferred into the OAIS.

But what happens with the records of no archival value? Some organizations might elect to store this data in some form of intermediate archive. This will make sense especially when no format conversion is needed. As soon as the long-term aspect becomes more relevant - including migration and changing access rights - the storage in the OAIS will be an option. This will have as a result that the OAIS stores some records for eternity and some only for a few years.<sup>20</sup> In this case the OAIS needs a deletion process that regulates very clearly who has deleted what and for which reasons. If records of archival value could be deleted from an OAIS this would be a serious breach of authenticity and would render many of the security measures useless.

## COORDINATION

For the Ingest process an agreement between the producer and the archives is necessary. In the traditional paper world this workflow must be coordinated in order to avoid 5 trucks of paper records arriving unexpected on one day. Such a process would create backlogs and confusion. But at least paper is patient and can be put in storage for a long time.

In the electronic world there must be a higher level of coordination. Producer and archive have to agree to the organizational and technical aspects of transferring

<sup>20</sup> Michael Hollmann, *Creating an intermediate archive for semi-active electronic records - concepts and challenges*: [http://www.dlm2008.com/img/pdf/hollmann\\_ab\\_gb.pdf](http://www.dlm2008.com/img/pdf/hollmann_ab_gb.pdf).



the records. The archive must know exactly which producer will transfer files to the archive and in exactly which format(s). As a first step the producers need to export the data from the original recordkeeping systems. In addition to this content a lot of metadata is needed to form a complete SIP. This SIP is then sent to a transfer platform.

Most of the metadata needed for the following processes must already be included in the SIP, like technical, legal and descriptive data, enabling storage, data management and access. Only when the metadata is complete the archive can start to convert the SIP to the archival storage units AIP.

## TRANSFER PLATFORM

The producers are responsible for placing the files on this transfer platform. As mentioned above this can only happen in a formally agreed process. The archive must know which producer has the right to transfer what kind of files. After the transfer the archive must be notified.

The transfer platform will always have a limited amount of space for the data it can hold. There must be rules how much can be transferred to it and what needs to be done with data on it. The archive must have the capacity to take the files from the platform and ingest them into the OIAS. What happens, if the archive does not manage to take in all the records? Then the platform will run full and the producers are blocked from transferring. Will they then be allowed to delete the records? In the Austrian province of Upper Austria all files being stored longer than 6 month on the transfer platform are to be deleted - putting pressure on the archive not to create a backlog.

This also relates to the amount of transfer at a time. In paper we are used to periodical accessions of material, maybe once a year or once in 5 years. Nobody would think of a weekly transfer of paper files from one producer. In the electronic world there are more possibilities since the transfer can be automated in different ways. It can be completely manual from the export to the conversion until ingest. In such a case similar accession intervals as in the paper world are desirable.

But the transfer can also be partly or completely automated. At a certain time a trigger in the recordkeeping system starts an export which builds the SIP. The transfer platform recognizes the SIP, checks its authenticity and completeness and creates the AIP which is transferred to the storage. In a highly organized archive even the descriptive metadata can be generated automatically and added without intervention into the archival hierarchy.<sup>21</sup> Humans would only intervene if one of the checks fails sending an error report. This however requires very standardized data on the side of the producer, a strict ingest process and a clear structured Data Management on the side of the archives.

In an automated process the size of one transfer bears no more importance. As an automated process does not require anybody's intervention it is possible to transfer just one or thousands of files per day. As a result today's notion of accessions where the custody of a certain number of files is transferred and administered becomes obsolete. It is replaced by a general agreement with the producer and the subsequent transfer of individual files.

---

<sup>21</sup> Niklaus Bütikofer, *Use Cases and requirements in Archives. Presentation 11. Nov 2007*: [http://nestor.sub.uni-goettingen.de/pi\\_2007/2007-11-09--buetikoer--archives.pdf](http://nestor.sub.uni-goettingen.de/pi_2007/2007-11-09--buetikoer--archives.pdf).

## GENERATE AIP (ARCHIVAL INFORMATION PACKAGE)

A defined process creates an AIP from a SIP. Their relationship can be 1:1 where each SIP is transformed to one AIP. This could be one record being exported from the recordkeeping system, transformed to a SIP which is then the basis for the AIP. But more often a SIP will consist of for example 5000 files from a recordkeeping system. There is now the choice to create one AIP with 5000 files or 5000 AIPs. The descriptive metadata for Data Management will of course always create 5000 units of description.

The best solution depends on three considerations: (1) The complexity to handle a great number of files, (2) the difficulty when migrating big files with a number of different formats and (3) most importantly on the form of access. In Access a user will search for this metadata and find one file of interest. After the request for this file the AIP will be transformed to a DIP and presented to the user. If this DIP consists of all the 5000 files the result is not very user friendly. When this user has the permission to see this one file but not the other 4999 then the DIP becomes unusable.

Therefore the size of the SIP, AIP and DIP has to be considered in the initial stage of a project. The easiest solution of course is when each information package has only one content object (like a file) but it also means the highest possible number of files being transferred and managed.<sup>22</sup>

## FINAL FORMAT

A very important decision during export and ingest is the question of the final format of files. While the technical formats for long-term preservation are a question answered by many projects and proven in reality I want to focus on the point in time when the files are converted in the agreed final formats.<sup>23</sup>

Many projects consider converting all files into a final format during the Ingest process. This has the advantage that all files are produced under the same controlled environment and the preservation metadata is generated to the same standard. On the other hand this choice has many disadvantages. Giving the archives the sole responsibility for the conversion process can overwhelm even larger organizations. Conversion can use large computing and time resources and always needs some form of checking. If the conversion process fails it puts all responsibilities on the archives.

Therefore the better solution is the conversion into the final format as soon as possible. This can be right after the creation of a document or at least after the last changes were made, like when a record is closed. This includes that the document is used and checked in the normal life-cycle of records making it easier to discover any errors in the conversion process. The archive receives most of the data already converted and checked making its workload more manageable.

---

<sup>22</sup> *The Danish National Archives for example have chosen this approach to make one SIP equal to one AIP. This is also possible because the legislation gives them the possibility to control the export and conversion process. The producers have to follow the archives instruction in creating the SIP: Jan Dalsten Sørensen, Revising the standards for digital transfers. From records management to archives - developing standards for transferring records: [http://www.dlm2008.com/img/pdf/sorensen\\_ab\\_gb.pdf](http://www.dlm2008.com/img/pdf/sorensen_ab_gb.pdf) .*

<sup>23</sup> *A very important initiative on preservation is currently Planets, Preservation and Long-term Access through Networked Services: <http://www.planets-project.eu>. See also InSPECT: <http://www.significantproperties.org.uk/index.html>.*

## AUTHENTICITY

During the export and the Ingest process the archive has to guarantee the authenticity of all the records. This means that certain checks are done and the logs for these checks are kept just as long as the records themselves.<sup>24</sup>

For the authenticity the archive must proof that the records were not changed, for example with a checksum.<sup>25</sup> Also the records must be readably - by the machine and by humans. While this seems to be a simple task it can be a challenge after several conversion steps. Lastly the records also must be understandable, which is more than just the ability to read something. For example if a MS-Word document containing a spreadsheet is converted into PDF the result might contain all the numbers of the spreadsheet but the lines and the spaces between the numbers might have been lost or at least distorted. In this case the numbers are readable but they make no more sense to the reader.

## DATA MANAGEMENT

During Ingest the descriptive metadata is duplicated as it is first stored in the AIP where it remains unchanged and authentic. Secondly it is stored in the Data Management for search and retrieval purposes. Here they are the basis for the units of description. Each record from the recordkeeping system must again be search- and viewable as a record. This metadata can be added and changed for quality assurance or more detailed description as it is usual in the archival tradition. Changing the metadata in the Data Management does not change the metadata in the AIP.

Furthermore the connection between records of the same series and the responding fonds over multiple transfers must be guaranteed. Therefore the Data Management must be able to create a hierarchy of description according to ISAD(G). Usually archival information systems (AIS) like scopeArchiv provide these functionalities.<sup>26</sup> An AIS also guarantees the connection between the traditional and the electronic archives. Both forms should be administered in the same hierarchy and with the same basic description. Different systems for different forms of data are an unacceptable solution for archivists.

The Data Management is responsible for storing and administering also the access rights providing an important legal service. Lastly the Data Management also provides the basis for the access of the consumers who search the metadata and are provided with the content.

The Data Management is the central tool for the internal administration and is closely connected to all other parts. It receives the metadata from the Ingest, provides the link to the AIPs in the Archival Storage, gives access to the consumers and is generally needed for all administrative functions of the OAIS.

---

<sup>24</sup> Lisa LaPlant, Kate Zwaard, *A holistic approach for establishing content authenticity and maintaining content integrity in a large OAIS repository*. Archiving 2008, June 24-27, 2008, Berne, Switzerland.

<sup>25</sup> Mariella Guercio, *Authenticity and OAIS. The CASPAR Model and the Interpares Principles and Outputs*. 11. 6. 2008: [www.digitalpreservationeurope.eu/preservation-training-materials/files/AuthenticityOAIS.ppt](http://www.digitalpreservationeurope.eu/preservation-training-materials/files/AuthenticityOAIS.ppt).

<sup>26</sup> *scope solutions ag*: [www.scope.ch](http://www.scope.ch).

## ARCHIVAL STORAGE

Archival storage seems to be the easiest part. We all know that disk space is very inexpensive so storage is considered cheap. All archival organizations have some kind of storage either in-house or in agreement with an IT-service. So the storage part does not seem to be a big challenge. When looking closely into the functions of this module some more daunting problems arise.

When it comes to implementing a system that involves checks, backup and most of all people who run this system it can be very expensive.<sup>27</sup>

Especially the “monitor technology” and “develop preservation strategies and standards” is something that needs a lot of expertise. This involves a close look at the existing technologies, the capability to see future trends and a team to develop strategies for migration or emulation. For me this sounds a lot like a job for IT-people rather than for archivists. They should however work closely together on these issues.

It also becomes obvious that this is not a job for small or mid-sized archives. Only those organizations with enough money and human resources can employ enough specialists.

They need to collect the relevant information and distribute it to the community. Today it is not clear who will provide this service; it could be one organization on a national level or even only a few on the international level.<sup>28</sup>

The same is true for the physical storage. While storage space is not an issue it is the storage services that can be quite difficult and expensive. Time will show whether storage will be done on a regional, national or again at an international level where for example 5 - 10 storage centers provide the service for all of Europe.<sup>29</sup>

## ACCESS

In Access the information is provided to the consumers. They are not a homogenous group but a diverse clientele with different interests and access rights. It is the task of the Access to provide only the information that somebody has the right to see but in a form that is useable. While the first part is guaranteed by the rights information stored in Data Management the second is provided by the DIP. The

<sup>27</sup> Ashley, Kevin (1999) “Digital Archive Costs: Facts and Fallacies.” In *Proceedings of the DLMForum on Electronic Records (DLM '99)*, p. 123: [http://europa.eu.int/ISPO/dlm/dlm99/dlm\\_proceed99\\_03.pdf](http://europa.eu.int/ISPO/dlm/dlm99/dlm_proceed99_03.pdf)  
 Stephen Chapman, *Counting the Costs of Digital Preservation: Is Repository Storage Affordable?*: [jodi.tamu.edu/Articles/v04/i02/Chapman/chapman-final.pdf](http://jodi.tamu.edu/Articles/v04/i02/Chapman/chapman-final.pdf).  
 Lance Watsky, *A Cost Comparison and Analysis of Storage Options for Media Collections*: [http://mic.loc.gov/preservationists\\_portal/presv\\_costcompare.htm](http://mic.loc.gov/preservationists_portal/presv_costcompare.htm).  
 Long-term Preservation Storage: OCLC Digital Archive versus Amazon S3 (May 16th, 2008): <http://dlj.org/article/oclc-digital-archive-vs-amazon-s3/>.

<sup>28</sup> For example the Digital Longevity project: The Dutch National Archive has established a Digital Longevity department to help ensure that Dutch government digital information is sustainable, properly managed, and can be preserved in an authentic and re-usable manner for the long-term. Also the UK National Archives provide a lot of information and services.

<sup>29</sup> The Austrian government is currently in the process of selecting an archival system that includes the storage. It is designed as a PPP (public-private-partnership) model, where national and international data creators will be invited to archive their data in the central storage: *Konzeption und Bereitstellung einer Software-Lösung für die digitale Langzeitarchivierung von Verwaltungsinformationen des Bundes (digLAimBUND) im Österreichischen Staatsarchiv*: <http://www.ausschreibung.at/smi-nr-114892-info-Konzeption-und-Bereitstellung-einer-Software-L%F6sung-f%FCr-die-digitale-Langzeitarchivierung-von-Verwaltungsinformationen-des-Bundes-digLAimBUND-im-%D6sterreichischen-Staatsarchiv.htm>.

DIP is created from the AIP in accordance to a set of rules. As mentioned above the relationship of AIP to DIP must not be 1:1.

The Access has the job to present the data readably and authentic. But it should also make sure that the information is understood by the users. This is true not only for the content itself but also for the context. A lot of institutions who undertake an electronic archives initiative look at the need of the current users of these records. While this is a prerequisite their view must also go beyond the first 10 years. The following questions should be answered:

How will anybody not connected to the creation process find this record? This reflects back to the creation of the SIP. Data that is not included in this step cannot be searched by the users.

How will anybody not familiar with the creating organizations find this record? This is a traditional function of archives where fonds and series provide the general structure. In some long-term preservation projects these aspects are ignored, especially the fonds. They are needed to provide a consistent access. For the consumer it is also very important that traditional and electronic records are kept in the same hierarchy with the same basic description. It cannot be the job of the user to know in which system records were created. One archival information system has to provide the access to all the data.

## **CONCLUSION**

The long-term preservation of electronic records does not start and is not finished with buying large storage systems. At the beginning there must be a clear concept of who the producers are, what they are producing and how the data is transferred to the archives. This concept must include the information that is needed for the preservation and for the dissemination to the consumers because most of that information is generated either by the producer or during the Ingest process.

As mentioned above in the chapter Archival Storage this task is a daunting one for each archives and might proof impossible for small and mid-sized archives. Bigger initiatives or archives must provide guidance; commercial companies must work as consultants. No single standard software has yet emerged and likely will not for the next years that covers all the functions of OAIS. To make it commercially viable it has to cover several bigger projects in several countries. But this should not hinder in making decisions right now.

There are different approaches when it comes to software. The first approach that has been chosen most of the times in recent years is the in-house development. In individual projects the requirements were defined and subsequent a solution programmed.

This approach has the clear advantage that 100% of the requirements are met - providing the project has adequate funding and is finished in time. It has however the huge disadvantage that a lot of the capacity of an institution is required for the specification, implementation and testing period. Especially the time and effort for testing is generally underestimated. Also in the long run a sizable effort is needed for updates making it commercially not very interesting.

Another disadvantage is the amount of know-how for such a project. When developing in-house the specification is also compiled in-house using in-house knowledge. The experience of all the other institutions with similar projects is not

considered - or at least not as an essential part of the design. Therefore this solution is rarely the best overall solution - both in costs and in functions.

Currently the most favored approach is to buy commercially available parts of the OAI where the interoperability between the systems is guaranteed. This combines the advantages of the in-house solution with those of well tested and established standard software. Standard software is readily available, tested and has the input of many clients.

In such a scenario the export from the recordkeeping system and the creation of the SIP might be programmed individually. The SIP is transferred to a commercially available product where the ingest is administered. The metadata is sent on to a standard archival information system (AIS) that has a built in access functionality. The AIP might be custom-built and is transferred to a standard archival storage system that includes mechanism for authenticity. The interface between the SIP and the Archival Storage is programmed individually.

Another possibility is to buy a complete system by a vendor covering all the functionalities. This approach requires a lot of preparation, planning and specification and a lot of money at the beginning. If successful it provides the archives with a single system that should meet all the requirements.

Large archives like the national archives of the United States of America or United Kingdom have been successful with this approach. They found vendors who custom-built individual systems. But it might be very difficult and costly for smaller institutions as "off the shelf" commercial software is not yet in sight.

The overall picture of long-term preservation including the organizational issues and the technical systems is quite complex. From the first steps of the initial design phase until after implementing all the parts a measure to ensure success is needed. To meet this challenge several initiatives respectively defined a catalog for "Trusted Digital Repositories". They provide more or less checklists for all the necessary aspects.<sup>30</sup> These documents provide an excellent guide on how to build an electronic archive.

---

<sup>30</sup> *Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report (May 2002):*  
<http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>.  
*OCLC, Audit Checklist for Certifying Digital Repositories:*  
<http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2433.html>.  
*nestor - materialien 8 - nestor-Kriterien - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version II (November 2008):* [http://www.langzeitarchivierung.de/downloads/mat/nestor\\_mat\\_08.pdf](http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_08.pdf).  
*Ronald Jantz, Michael J. Giarlo, Digital Preservation. Architecture and Technology for Trusted Digital Repositories. D-Lib Magazine (June 2005). Volume 11 Number 6:*  
<http://www.dlib.org/dlib/june05/jantz/06jantz.html>.  
*Reagan W. Moore, MacKenzie Smith, Assessment of RLG Trusted Digital Repository. Requirements:*  
[http://sil.unc.edu/events/2006jcdl/digitalcuration/Moore\\_Smith-JCDLWorkshop2006.pdf](http://sil.unc.edu/events/2006jcdl/digitalcuration/Moore_Smith-JCDLWorkshop2006.pdf).  
*Christopher A. Lee and Helen R. Tibbo, Digital Curation and Trusted Repositories: Steps Toward Success. Journal of Digital Information, Vol 8, No 2 (2007):* <http://journals.tdl.org/jodi/article/view/229/183>  
*Data Seal of approval:* <http://www.datasealofapproval.org>.

## POVZETEK

### ARHIVIRANJE ELEKTRONSKIH DOKUMENTOV V SKLADI Z MODELOM OAIS - ORGANIZACIJSKI IZZIV

Dandanes predstavlja na področju arhiviranja največji izziv arhiviranje in dolgoročna hramba dokumentov v elektronski obliki. Arhivska skupnost se tega problema zaveda, obstajajo zamisli za rešitev in je trenutno na poti iskanja ter vpeljevanja rešitev.

Najdemo lahko le nekaj »dobrih praks« o tem, kako se spopadati z izzivom arhiviranja elektronskih zapisov. Na splošno je bil sprejet le standard OAIS. To je referenčni model za dolgoročno hrambo. Opisuje več funkcij med arhivom, njegovim upravljanjem in ustvarjalci kot tudi z uporabniki dokumentov. V članku želi avtor podati pregled modela OAIS in pokazati na nekatere izzive, s katerimi se pri njegovi vpeljavi srečujejo arhivi. Članek ne obravnava tehničnih vprašanj ali oblik datotek, osredotoči se bolj na organizacijske težave. Na več mestih avtor opozori tudi na nekatere najbolj pomembne projekte, ki nudijo podporo in odgovore.

OAIS definira ustvarjalce, in sicer so institucije ali osebe, ki arhivsko gradivo predajajo arhivom. Ustvarjalec ima dolžnost, da napravi izvoz iz originalnega sistema in ustvari informacijski paket, imenovan SIP (oddajni informacijski paket), ki ga sprejme arhiv. Ta potem izvrši številne kontrole paketa, ki mora biti nato preoblikovan in shranjen kot AIP (arhivski informacijski paket). Ko so podatki dokončni in pripravljeni za arhiviranje, se jih dalje obdeluje. Najprej se v »Data Management« shranijo samo opisni metapodatki, AIP pa je shranjen v arhivski hrambi.

V »Data Management« se shranjujejo vsi opisni podatki o elektronskih dokumentih; to se najverjetneje zgodi v arhivskem informacijskem sistemu (AIS). V arhivski hrambi (Archival Storage) so podatki zaščiteni in hranjeni v sistemu. Tukaj se tudi izvaja preverjanje napak, obnavljanje in prenos podatkov. Dostop (Access) vsem strankam omogoča možnost iskanja in pridobivanja arhiviranih podatkov. Podatki se uporabnikom predajajo kot DIP (predajni informacijski paket). Kot dodatek opisanim funkcijam obstaja še administracija, kjer se pripravljajo dogovori glede predaje in prevzema SIP, pravila dostopa in kjer se obravnavajo splošni standardi in politike. Načrtovanje varovanja podatkov kontrolira tehnološke spremembe in razvija strategijo, ki vodi k prenosu podatkov. Pomembno je vedeti, da so mnoge od zahtev dolgoročne hrambe in varovanja elektronskih zapisov, organizacijske potrebe. Šele takrat, ko so te izpolnjene, se lahko začne implementacija tehničnih sistemov.

To velja še posebej za proces prevzema, kjer se podatki od ustvarjalca prenašajo v arhiv. V svetu obstaja že nekaj projektov in smernic, ki nudijo pomoč pri vzpostavljanju podrobnega in natančnega procesa. Ta del zahteva veliko koordinacije v pravnem, organizacijskem in tehničnem pogledu. Vključuje valorizacijske odločitve in čas, ko se valorizacija izvede, ne glede na to, ali se izvede pred, med ali po prevzemu. Ustvarjalci in arhivi se morajo dogovoriti, kakšni metapodatki bodo vključeni v proces. Za kasnejši dostop so neprecenljivi podatki o vsebini in dostopnih rokih. Prav tako morajo uskladiti čas in količino prenosa, saj morajo arhivi priskrbeti potrebne vire za prevzem vseh podatkov. Dogovoriti se je potrebno tudi o končnih formatih in verodostojnosti podatkov.

V »Data Managementu« se za iskanje in pridobivanje podatkov morajo shraniti tudi opisni podatki. Za zagotavljanje kakovosti in bolj podrobnih podatkov, kot je to

običajno v arhivistiki, se ti lahko dodajajo in spreminjajo v arhivskem informacijskem sistemu (AIS). Tukaj se po standardu ISAD(G) hrani hierarhija fondov, serij, združenih dokumentov in dokumentov. AIS prav tako zagotavlja povezavo med klasičnim in elektronskim arhivom. Obe obliki moramo upravljati v isti hierarhiji in z istim osnovnim popisom.

Različni sistemi za različne oblike so za arhiviste nesprejemljiva rešitev. Del sistema, ki predstavlja arhivsko hrambo, se zdi najbolj enostaven, vendar ko pride do vzpostavitve sistema, ki vključuje kontrolo, varnostne kopije in ljudi, ki ta sistem nadzorujejo, je lahko zelo drag. Funkciji »nadzor tehnologije« in »razvoj strategije varstva in standardov« zahtevata veliko znanja in zanju je potrebnega veliko časa. Očitno je, da to ni najbolj primerna naloga za majhne ali srednje velike arhive. Dandanes še ni jasno, kdo bo te storitve ponujal; lahko pa so izvedene na regionalnem, državnem ali celo na mednarodnem nivoju.

Zelo pomembno vprašanje za vse projekte so informacijski paketi - torej SIP, ki so poslani od ustvarjalca v arhiv in AIP, ki hranijo informacije in končno še DIP, ki so ustvarjeni za uporabo. V članku avtor podaja pregled najbolj pomembnih standardov na tem področju in kako jih lahko implementiramo. Dolgoročna hramba in zaščita elektronskih zapisov se ne začne in konča samo z nakupim velikih sistemov za hrambo. Na začetku moramo imeti jasn koncept o ustvarjalcih, kakšno gradivo ustvarjajo in kako se podatki prenašajo v arhiv. Ta koncept mora vključevati informacije, potrebne za zaščito in končno razširjanje gradiva uporabnikom, saj se večina teh informacij generira pri ustvarjalcu ali med procesom prevzema.

Manjši arhivi morajo delati skupaj ali pa se zanašati na rezultate večjih iniciativ. Potrebne nasvete lahko dajo tudi svetovalci. Ko so dogovori o konceptu končani, se začne iskanje in implementacija sistema. Ker ne obstajajo produkti »s police«, ne moremo pričakovati enostavnih rešitev. Najbolje je najprej poiskati najpomembnejše komponente in potem videti, če in kako te delajo skupaj - pri tem pa vedno imeti v mislih celoten koncept. Celotna slika dolgoročne hrambe in zaščite, ki vključuje tudi organizacijska vprašanja in tehnične sisteme, je zelo kompleksna. Od prvih korakov do vpeljave vseh delov so potrebna merila, ki jamčijo uspeh. Za spopad s tem izzivom je nekaj iniciativ definiralo katalog za »Zaupanja vredna digitalna skladišča«, ki podaja kontrolne sezname za vse potrebne vidike. Ti dokumenti predstavljajo odličen vodič o tem, kako zgraditi elektronski arhiv.