# DIGITAL PRESERVATION FOR ENTERPRISE CONTENT: A GAP-ANALYSIS BETWEEN ECM AND OAIS

**Joachim Korb**
AIT Austrian Institute of Technology
Safety & Security Department

**Stephan Strodl**
Vienna University of Technology
Information and Software Engineering Group

## ABSTRACT

Today, more and more information is being produced in a digital form. In addition to this so-called born-digital content, material that was produced to exist in an analogue form is now being digitised both for preservation and for easier access. This digital information comes in an ever greater variety of formats, many of which are relatively short-lived. Newer versions of the same software are often unable to render files produced with older versions of that software, let alone files produced with similar software from other vendors. Soft- and hardware environments change constantly and after only a few years can older files often no longer be rendered with up-to-date systems.

While large scientific organisations and memory institutions (museums, libraries and archives) have in recent years invested significant effort and activity towards digital preservation, the commercial world does not currently have the means to preserve their digital information for the long term.

This paper sets out to determine what would be needed to make modern Enterprise Content Management (ECM) Systems ready for long-term preservation of the assets stored within them. For this aim both the general Model of an ECM and the "Reference Model for an Open Archival Information System (OAIS)" have been described, and the special needs of an enterprise system identified.

A special focus lies on the Electronic Records Management (ERM) component of ECMs, which already provides simple preservation functionalities, but lacks those aspects of the OAIS that would make it truly long-term preservation capable. A truly long-term preservation capable ERM would have to add these while retaining capabilities of compliance (the retention or destruction of certain documents in accordance to legal requirements).

## 1. INTRODUCTION

Over the last decades more and more information has been produced in a digital form. While at first computers may only have acted as 'intelligent typewriters', an increasing part of what used to exist only in an analogue form is now held digitally. Additionally, material that was produced to exist in an analogue form is now being digitised both for preservation and for the purpose of wider access. This digital information comes in an ever greater variety of formats, many of which are relatively short-lived. Oft-cited examples for this are the Microsoft Word format, which has changed continuously over the different versions of the software, and CAD files which are so reliant on the software they were produced with that it is usually impossible to render older files with newer versions of the software, let alone software from a different vendor.

As time progresses soft- and even hardware environments change, so that after only a few years it is often no longer possible to open older files. To understand the problem, two aspects of digital preservation must be considered. First, there is the question of preserving access to the actual bits of digital information; this is usually referred to as bit-stream preservation. Bit-stream preservation includes questions of media integrity as well as the hardware necessary to read the media.

This aspect of the problem is already solved in many ECM solutions, but, while keeping redundant copies of all data may safeguard against the loss of the actual files, it will not guarantee their long-term[1] accessibility. The danger of losing whole collections of data as a result of outdated data formats, software or run-time environments must be countered by developing Digital Preservation Systems – this aspect is known as logical preservation. Logical preservation is the main concern of this gap-analysis.

Large scientific organisations and memory institutions have in recent years invested significant effort into ensuring the long-term availability of their entrusted digital assets. For this, both commercial vendors as well as a number of smaller and larger projects[2] have produced systems for storing, managing and accessing those assets. A special focus in all these efforts has been on the compliance with the "Reference Model for an Open Archival Information System (OAIS)" [4]. An ISO-standard since 2003, this reference model not only describes a system long-term

[1] [4] defines "long term" as "a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats […] on the information being held in a repository. This period extends into the indefinite future".
[2] E.g. the EU-funded PLANETS project [http://www.planets-project.eu/] or the Austrian RS-DME project [http://www.rs-dme.at/], for which the original version of this analysis was written.

preservation but provides a common vocabulary to those concerned with such work.

In the commercial world the situation is quite similar to that described above. Large volumes of born-digital and digitised material are ingested into, managed in and accessed from what is now commonly called Enterprise Content Management (ECM) Systems. The main difference between these and the systems in use in many of the above mentioned institutions is that ECM Systems do not, currently, provide the means to keep the stored information accessible in the long term.

The aim of this paper is to determine what would be needed to make modern ECMs ready for long-term preservation of the assets stored within them. For this aim both types of systems will be compared, the special needs of an enterprise system will be identified and the steps to make a typical ECM OAIS-compliant will be described.

In the enterprise environment there are a number of different terms describing systems used to store digital information. This is mainly due to two facts:

1. These terms were coined to advertise software; and different companies would sell their products under a variety of names to differentiate between theirs and similar products from other companies, but also from earlier versions of their own products that had fewer or different capabilities.
2. Until recently there has been no common model for these systems.

For a long time, the term Content Management was used to describe systems that allowed enterprises to manage document and content flow. Now, however the term Content Management System is most often used to describe software for maintaining, controlling, changing and reassembling the content for internet presentation.

For the purpose of the gap analysis, this paper will follow the Association for Information and Image Management's (AIIM)[3] definitions [2] and use the term Enterprise Content Management System, to describe the strategies, methods and tools used to manage business content. For the description of a long-term preservation archive the Consultative Committee for Space Data Systems' (CCSDS)[4] has produced the OAIS Reference Model. In this analysis, the OAIS terms and definitions will be used when referring to such a system.

The remainder of this paper is organised as follows: Section 2 presents the model of an Enterprise Content Management, followed by the description of the key concepts of the OAIS reference model in Section 3. The results of the gab analysis are presented in Section 4.

[3] http://www.aiim.org/
[4] http://public.ccsds.org/

## 2. ENTERPRISE CONTENT MANAGEMENT

The following description of a model for Enterprise Content Management follows the description of AIIM. In 2005, the Association for Information and Image Management, "the leading non-profit organization focused on helping users to understand the challenges associated with managing documents, content, records, and business processes,"[5] set out to find a common name and description to identify the procedures as well as the types of systems that allow for the control of enterprise content.

AIIM describes ECM as "the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes."[2] These tools, methods and strategies are used to manage what is referred to as "the lifecycle" of that content.

For the purpose of this description, the model used here will follow AIIM's descriptions rather than its images. The problem with the images AIIM uses to describe its model is that they are designed to reflect a rather complex concept in a way that makes them perfect for marketing. The complexity of the model is due to the fact that it encompasses strategies, methods and tool, while trying to leave enough room both for the description of different settings and for vendors to emphasise the respective merits of their own software.

### 2.1. ECM Components

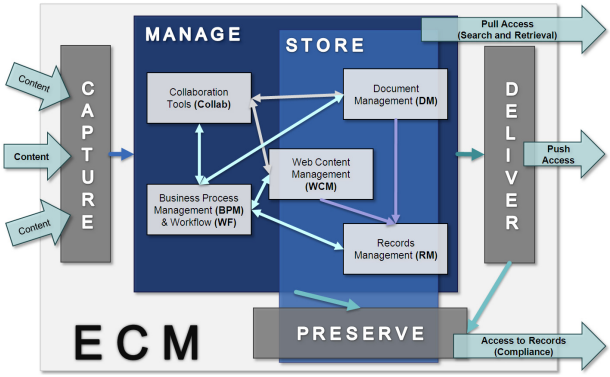


**Figure 1**. ECM Components Model

The ECM Components model (Figure 1) shows the main components of an ECM, which are:

**Capture (Input Management)**: Capture's function is to ensure that any content within a certain setting is managed by the ECM system as soon as it is created. An important part of Capture is the automatic classification of that content.

**Manage**: Manage is the component that provides the management of each individual content item through all

[5] http://www.aiim.org/AboutAIIM/ECM-ERM-BPM-Association.aspx

its versions (lifecycle). It includes the following management applications:

- **Document Management (DM)**: A document in the context of ECM Document Management is defined as recorded information or an object which can be treated as a unit [1]. Today, this also includes E-Mail Management (EMM) and Digital Assets Management (DAM)[6].
- **Collaboration Tools (Collab)**: Collab includes the joint use and control over the content (including access management), as well as the applications that support this.
- **Web Content Management (WCM)**: Web Content Management is often controlled through what is now called a Content Management System (CMS) and may or may not be directly integrated with the ECM. Many ECM solutions provide access to the content via web-based user interfaces and many include actual WCM functionality.
- **Records Management (RM)**: Records Management is the management of what in this context is called Records. Records are content which will not change further and which, for legal reasons or because it may be of further relevance to the enterprise, must be stored for future reference [8]. In enterprise as well as in government environments, (electronic) records management (ERM) is governed by ISO standard 15489 [5].
- **Business Process Management (BPM) / Workflow (WF)**: BPM is a methodology to make processes efficient and effective by developing, deploying, monitoring, and optimising process automation applications. WF, as opposed to BPM, is the manual processes of managing documents in cases where human intervention is required (e.g. approval and prioritisation).

**Store**: The Store component of an ECM includes the actual physical locations (e.g. hard disks, storage area networks (SANs), or even CDs/DVDs) where the content is stored, as well as the logical structure of these physical locations. That structure, referred to as 'repository', can be a simple file system, a database or even a 'data warehouse'. Store also includes access strategies, also called 'library services', which include controlled check-in and check-out of content, search and retrieval mechanisms, version control, and the audit trail of each individual item. As such, it has a significant overlap with Manage.

**Deliver (Output management)**: As search and retrieval (pull access) strategies are already controlled in the Manage and Store components of this model, Deliver is not concerned with this aspect. Deliver focuses on the control of external access to, and publication and distribution (push access) of content. This includes transformation of content for external access (e.g. text documents into personalised serial e-mails or letters or into PDFs for web publication), but also compression of files for storage or transformation of e.g. text documents to PDF/A [6] files for Preserve.

**Preserve**: The Preserve component deals chiefly with content that has been identified as Records by the Manage component. It is, obviously, directly related to the Store component as it deals with safe, long-term storage and back-up strategies (bit-stream preservation) for these Records. Preserve is fed either directly by Records Management and Manage or indirectly via Deliver, when content has been transformed for archiving.

### 2.2. ECM Compliance, Records Management, Preserve and Long-term Archiving

There are two main reasons why enterprise content that is no longer in regular use is preserved:

The first is known as Compliance. According to AIIM Compliance "means ensuring that the proper business practices are followed and that content is properly captured, stored, managed, and disposed of at the appropriate and legal time in its lifecycle."[3] This lifecycle may last for 10 years or more, during which time the respective content may need to be accessed, destroyed[7] or passed on to a different organisation (e.g. a national archive) at request or at a given time. It may be important that the business is able to prove the proper and legal destruction of said content, which will only be possible if that content is still accessible at the time it is to be destroyed. The second reason is more directly business related: Content may contain information about previous developments or projects, and this information may be of importance to later developments or projects.

Content that is no longer in active use but kept for either of the above mentioned reasons is referred to as Records. There is an ISO standard that regulates the procedures in handling Records. According to ISO 15489, a Record is "information created, received, and maintained as evidence and information by an organisation or person, in pursuance of legal obligations or in the transaction of business" and (Electronic) Records Management ((E)RM) is the "field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including processes for capturing and maintaining evidence of and information about

[6] DAM deals with any digital content that cannot be classified as a document.

[7] This is an important point in the comparison as the OAIS Model does not support the destruction of content.

business activities and transactions in the form of records."[5]

As has been said above, Records are found in two of the main ECM components. The Manage component is responsible for deciding which Content will be kept as Records, while the Preserve component is responsible for actually archiving these Records. It is important to note that this mainly involves the storage of the Records. Migration in this context is still largely the migration of data from one storage medium to another and not, generally, the migration of content to different file-formats when the original format becomes obsolete. Content will only be migrated when passes through Deliver. This happens rather as a general strategy than as an actual act of digital preservation. If the format is chosen well (e.g. PDF/A for text documents) this may, however, have a similar effect for the accessibility of the content.

AIIM describes their approximation of Long-Term Archiving as "content that must be preserved over decades must be saved to media, such as paper and film-based imaging, with longevity to match."[2] There is some discussion about the 'transformation' of content into file formats that are preferable for long-term preservation. AIIM itself takes part in the development of the PDF/A standard. There are also considerations in the ECM and the ERM community about the transformation of e.g. CAD files into TIFF, JPEG or JPEG2000 files. [9] describes an example to keep CAD files usable, i.e. re-usable in a later projects by migrating them an active CAD file-format, as it is impossible to produce e.g. a new architectural plan from static image files. These transformations are handled by Deliver and are thus connected to that part of the ECM system which usually deals with push-access and delivery of content, not with its preservation. No mention is made of management of long-term preservation that includes actual preservation planning as envisioned in the OAIS Model.[8]

## 3. THE OAIS MODEL

The OAIS Reference Model was first published in 1995, when the partners of the Consultative Committee for Space Data Systems realised that large portions of their data were no longer accessible due to changes in software and hardware systems. The model has been continuously refined. In January 2002 the OAIS Reference Model was published as a CCSDS Blue Book and has subsequently been adopted as an ISO Standard (ISO 14721:2003).

Reference Model standards, like OAIS, are developed in an open, public process. As the problems this standard addresses have become important beyond the space communities, the CCSDS set out to ensure broad participation from other fields – notably from the traditional archive community.[9] Since space data was no longer the only subject of the resulting model, the term Information was used to identify the content that would then be represented by the data in the archive.

### 3.1. Description

In the OAIS Reference Model, the Archival Information System includes hardware and software components as well as the people who are responsible for the acquisition, preservation and dissemination of the information. Additionally, the Model is designed as a framework for understanding, applying and discussing concepts needed for long-term digital preservation of information. Long-term, in this framework, means "long enough to be concerned about changing technologies."

#### *3.1.1. OAIS Information and Information Package Definitions*

One of the most important concepts in the OAIS Reference Model is that of **Information**. Information is defined as "any type of knowledge that can be exchanged." This Information is always represented as **Data**; and each individual instance of such Information is identified as an **Information Object**.

In order for an Information Object to be successfully preserved, it is necessary for the OAIS to clearly identify and understand the **Data Object** (the Data associated with that instance), and its associated **Representation Information**. The Representation Information is additional information that maps a Data Object into more meaningful concepts. Only in this combination is the Data Object usable and becomes the Information Object (or the object that was to be preserved). Without the Representation Information, the Data Object is often useless (see Figure 2).

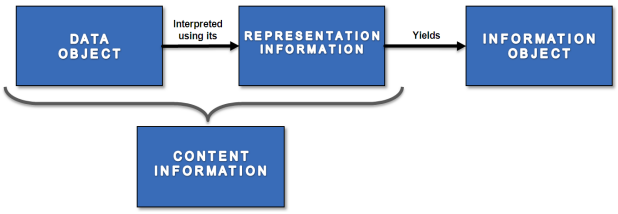


**Figure 2**. Relationship between Data Object and Information Object

Closely related to the concept of Information is that of the Information Package. An Information Package is a conceptual container of two types of Information called **Content Information** (the combination of Data

[8] Kampffmeyer mentions the OAIS Model in [7] as a standard related to migration, but does not, apparently, consider the importance of preservation planning for ERM.

[9] The "Open" in OAIS is meant to signify this aspect of the modelling process and does not imply Open Access to the OAIS's content.

Object and its associated Representation Information) and **Preservation Description Information**.

Preservation Description Information is the Information which is necessary for adequate preservation of the Content Information. It contains the following information:

**Provenance**: Provenance describes the history of the Content Information: Where it originated, what was changed (e.g. necessary format changes), who had custody of it since creation.

**Reference**: Reference identifies the Content Information (similar to an ISBN for a book).

**Fixity**: Fixity provides the authentication mechanisms and authentication keys to ensure that the Information Object has not been altered in an undocumented manner. This function is closely related to the concept of archival Authenticity, which is also relevant in Records Management and Compliance.

**Context Information**: Context Information documents the relationships of the Content Information to its environment.

### *3.1.2. OAIS Roles*

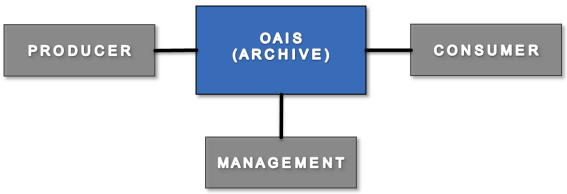


**Figure 3**. Simple OAIS Model

The simplest view of an OAIS (as shown in Figure 3) has three major roles attached to it:

**Producer**: Producer is the role of the entities (persons or client systems) that provide the Information to be preserved in the OAIS.

**Management**: Management is the role of those entities that set overall OAIS policy. These will usually have further management functions in the organisation the OAIS belongs to.

**Consumer**: Consumer is the role of those entities (persons or client systems) that interact with OAIS services to search for and access preserved Information.

An important OAIS concept related to the Consumer is that of the **Designated Community**. This is an identified group of potential Consumers of the OAIS. The Information to be preserved should be **Independently Understandable**. This means that it must be documented in such a way that any member of the Designated Community can understand it without external resources. The need to achieve this informs the decision on the content of the Representation Information. The broader the future community of potential Consumers is to be, the broader the content of the Representation Information must be.

### *3.1.3. OAIS Functional Entities and Data Flow*

The following is an explanation of the functional entities of an OAIS (shown in Figure 4) and follows the flow of Data through the OAIS:

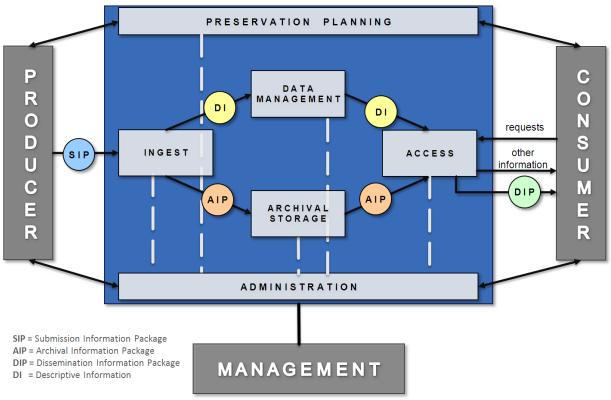


**Figure 4**: OAIS Functional Entities

**Ingest**: Ingest provides the services and functions for the OAIS to accept Information Packages from the Producers. These Packages are called **Submission Information Packages** (SIPs). The delivery of a SIP is negotiated between the Producer and the OAIS. At this point the OAIS assumes sufficient control of the Information to ensure long-term preservation. This means that it reserves for itself the right to manipulate the SIPs in such a way that they can be preserved. The OAIS also ensures that the Information is Independently Understandable to Designated Community by associating adequate Representation Information to it at this point. The OAIS follows documented policies and procedures (Preservation Plans), which ensure that the Information is preserved against all reasonable contingencies (e.g. through migration to another format). At Ingest, the content of the SIP is prepared for storage and management within the archive. Preservation Description Information is added to the Information Packages. The resulting **Archival Information Packages** (AIPs) are transferred to Archival Storage. The associated **Descriptive Information** (DI), which consists primarily of package descriptions, is provided to Data Management to support Access (the finding, ordering, and retrieving of OAIS Information holdings by Consumers).

**Archival Storage**: Archival Storage provides the services and functions for the storage, maintenance and retrieval of AIPs. It organises refreshing of storage media in order to provide the capability to reproduce the archive holdings over time (*bit-stream preservation*). For disaster recovery, Archival Storage provides a mechanism for producing duplicate copies of the AIPs in the archive collection. Finally, Archival Storage provides copies of stored AIPs to Access.

**Data Management**: Data Management is the entity which provides services and functions for populating, maintaining, and accessing both DI and internal archive administrative data. It receives query requests from Access and generates result sets that are transmitted back to requesting Consumers. If the requested Data is available, Data Management generates a dissemination request which is sent to Access.

**Access**: Access supports Consumers in determining the existence, description, location and availability of Information stored in the OAIS and allows them to request **Dissemination Information Packages** (DIPs). A DIP is derived from part or all of one or more AIPs and is the Information Package that is sent to a Consumer. Among Access' functions are the finding aids, tools that provide an overview of the Information available in the OAIS.

**Administration**: Administration is the entity that manages the overall operation of the OAIS. Administration negotiates submission agreements with the Producers, manages system configuration, and develops the standards and policies for the OAIS. These include format standards, documentation standards, and the procedures to be followed during Ingest as well as the policies for storage management. Administration is responsible for Preservation Planning and for the audit of AIPs. The audit process must verify that the quality of the Data meets the requirements of the archive.

**Preservation Planning**: Preservation Planning is an important task of Administration. Preservation Planning interacts with Consumers and Producers to monitor changes in their respective service requirements and available technologies. Such requirements may include data formats, media choices, preferences for software packages or computing platforms, and available mechanisms for communicating with the OAIS (e.g. new finding aids for Consumers or ftp-up-load rather than SIP delivery by optical media for Producers). Preservation Planning is also responsible for tracking emerging digital technologies, information standards, and computing platforms (i.e. hardware and software), to identify technologies which could cause obsolescence in the OAIS's computing environment and thus loss of access to certain parts of the archive's holdings.

Internally, Preservation Planning develops packaging designs and detailed migration plans in order to implement Administration policies and directives. Preservation Planning receives approved standards and migration goals from Administration and implements these. Migration goals usually involve transformations of AIPs including, at times, transformations of the Content. Once the migration plan, associated AIP designs, and software have been tested and approved, the entire migration package is sent to Administration, which will execute the actual migration.

It is important to note that migration is not the only way to mitigate technology obsolescence. Other options include the emulation of obsolete hard- and software environments.

## 4. GAP ANALYSIS

From the descriptions in the previous sections can be seen that ECMs and OAISs have similar requirements in many areas. Other areas are only present either in one or in the other kind of system. One main difference, however, is organisational. While OAISs are conceived to be external organisations[10] that are independent of the creation process of the information they contain, ECMs actually facilitate such creation and control the whole lifecycle of the content. Thus, there is the Capture process that is designed to draw all creation of Information (or Content) of an enterprise into the ECM as a central point and as a first step to manage further versions and variations of that Information from within the ECM. Only when that creation process is over does the Content turn into Records and is handed over to ERM and Preserve. Most of the functionality that both ERM/Preserve and OAIS need (e.g. search and access) is provided within the ECM via Manage and Store. Certain provisions that are central to the OAIS Ingest function are also handled in an ECM system. An example of this is the creation of Metadata and Descriptive Information, which is already provided with the capture process and is maintained over the content's lifecycle.

Some ERM requirements are quite similar to those many archives or scientific organisations have. Archives live by rules that are not much different from those referred to by the term Compliance. In fact, the "Model Requirements Specification for the Management of Electronic Records" (MoReq2[11]) was produced at the request of the DLM Forum[12], which is an independent European community of both public archives and other organisations which deal in archiving, and records and information management. The need of a business organisation to preserve its internal information for future reference, on the other hand, may be likened to that of large scientific organisations, e.g. the CCSDS.

Organisations like AIIM have only recently recognised the fact that digital files have the inherent risk of becoming obsolete. This missing awareness may have been due to the relatively smaller size of the relative organisations' data archives, or to the differences in the length time that is usually envisioned for data preservation. Only in recent years have people in the field become concerned with more than just the hardware side of preservation. AIIM is now a partner in

[10] This can also be independent parts of the organisations that produce the information.

[11] The MoReq2 (http://www.moreq2.eu/) specification has been prepared for the European Commission with funding from the IDABC (http://ec.europa.eu/idabc/) programme.

[12] http://www.dlmforum.eu/. The acronym "DLM" means "Document Lifecycle Management".

the creation of the PDF/A standard, and several people in the field suggest image-file formats like TIFF, JPEG and more recently JPEG2000 for preservation. This can, however, only be a first step towards the integration of logical preservation into the ERM field.

As the number of formats used increases regularly, and more and more information is contained in a mix of formats (e.g. entire web-sites that contain sound and movie files besides the image and text files covered by the above mentioned suggestions), it is important that a culture of long-term digital preservation arises.

Such a culture has existed for a number of years in the scientific and the cultural memory fields, and the OAIS Model is its expression.

### 4.1. What is required to make an ECM OAIS compliant?

As has been described above, an actual OAIS is made up of the following functional entities: Ingest, Archival Storage, Data Management, Access, Administration, and Preservation Planning. The Access functionality is already provided in the ECM through Capture, and different parts of Manage and Deliver, as is some of the functionality of Ingest, Archival Storage, and Data Management. What is totally missing is the combination of Administration and Preservation Planning. Together, these set the framework for proper digital preservation. They also provide control mechanisms and standards, and prescribe preservation policies, which set rules about when a certain type of preservation action is to be used on an endangered format. Preservation Planning provides the technology watch function (usually via external databases or technical registries[13]), which provides Administration with the triggers for such preservation actions. If, for example the external registry indicates that a particular format is at risk of no longer being supported by any software, Administration uses that indication to determine that now is the right time to migrate (in ECM terms transform) all files of that format. In this case, it is also the Preservation Planning component's responsibility to provide the plan on how (e.g. with which software, using which parameters etc.) and to which new format the files are to be migrated.

One important part of the OAIS Model is the description of the data flow within an OAIS. This not only ascribes responsibility to OAIS functionalities for the content at different stages of that data flow, but defines the additional information the OAIS must provide in addition to the original content. As has been shown above, ECMs do have similar functionality that provides e.g. version control and originator information (Provenance in OAIS terms) for content, but once the content turns into Records, no further - and more specifically no preservation-related – information is added. This means that large parts of the Preservation Description Information are missing.

In an OAIS, Representation Information and Preservation Description Information are added to the SIP at Ingest. Part of this is the identification of the proper file formats (i.e. with which version of a program was the file created, is it really the type of file the extension (e.g. '.doc') indicates, etc.). This again is usually handled by external services.[14] These steps are important, as only properly identified files can be successfully migrated according to the regulations of Preservation Planning. At present, no such information is added to the ECM Record.

Some of the information added to the SIP is provided to make its content Independently Understandable to the Designated Community. It can be assumed that for many ECMs the Designated Community's knowledge is equal to that of the Provider. It can perhaps be argued that in these cases no such information needs to be added, but a specific analysis of the circumstances may be advisable. Potentially, properly collected Preservation Description Information may be used as a source for Knowledge Management, thus ensuring that knowledge about certain processes and content are not lost to an enterprise when important employees leave or processes change.

As has been said earlier, Store, Manage and Deliver in ECMs provide much of the functionality that Archival Storage, Data Management and Access do in OAIS. It is, however, important that the manner in which these latter components work together is different. After Ingest, the AIP and its corresponding DI are sent to Archival Storage and Data Management respectively. Both are refreshed for each applied migration action. Following a request from Access, parts of the AIP and its DI are recombined to form the DIP. This functionality is, of course, not present in ECMs.

Additionally ECM Manage fulfils certain requirements that are part of an OAIS's Administration function. However, all requirements that deal with the audit of AIPs and all aspects of standard and policy development are missing.

So, while certain components of ECMs fulfil requirements of OAIS components, none of them completely do so, because the main purpose of an ECM is to provide an environment that allows for the active manipulation of content while the main purpose of an OAIS is to preserve Records (the ingested Information). The addition of such OAIS functionality to Records Management and Preserve can ensure that the information in the Records stays as close to the ingested

[13] E.g. the PRONOM technical registry. (http://www.nationalarchives.gov.uk/pronom/)

[14] E.g. the JSTOR/Harvard Object Validation Environment (JHOVE). (http://hul.harvard.edu/jhove/).

original as is possible and that additional information is kept to provide authenticity to the original content.

It appears therefore advisable to turn ERMs into proper OAISs while using functionality that already exists in other parts of the ECM.

One important function of ERMs that is not foreseen in the OAIS Model is that of the – legally required – destruction of certain Records after a given period of time. This function would of course have to be implemented into the new system.

## 5. SUMMARY

In this paper we have described and compared a general model for Enterprise Content Management systems and the Reference Model for an Open Archival Information System in order to determine what would be needed to make an ECM system OAIS-compliant and thus long-term preservation ready.

We have seen that some of the functionality needed for an OAIS can already be found in ECMs, but much of what it needs to become truly long-term reliable is missing:

1) Where ECM-Capture collects all content an organisation produces, OAIS-Ingest needs to be provided with the information that is to be preserved.
2) ECMs are usually integrated into the organisational infrastructure whereas OAISes are often external organisations that take responsibility for the preservation of the information other organisations have produced.
3) Capture collects metadata about ownership, access rights, and other information needed for the active part of a document's lifecycle. Ingest, while also responsible for some of these, specialises in preservation related metadata: e.g. file formats, representation and preservation metadata.
4) In an OAIS, descriptive information is held separately from the actual data (which represents the information that is to be preserved).
5) ECMs provide no Preservation Planning, watch functionality or controlled continuous logical preservation. To the OAIS these are provided by Administration.

One important point that can be seen from this analysis is that ECM and OAIS do not contradict each other; the OAIS functionality supplements that of the ECM. It offers missing functionality which may become crucial in providing businesses with Content Management Systems that assure Compliance for their digital assets.

## 6. REFERENCES


[1] Association for Information and Image Management (AIIM) *What is Document Management (DMS)?* AIIM. http://www.aiim.org/What-is-Document-Management-Systems-DMS.aspx. (Last accessed on 2010-03-23)

[2] Association for Information and Image Management (AIIM) *What is ECM?* AIIM. http://www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx. (Last accessed on 2010-03-23)

[3] Association for Information and Image Management (AIIM) *What is Electronic Records Management?* AIIM. http://www.aiim.org/What-is-ERM-Electronic-Records-Management.aspx. (Last accessed on 2010-03-23)

[4] Consultative Committee for Space Data Systems (CCSDS) *Reference Model for an Open Archival Information System (OAIS)*. CCSDS, 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf. (Last accessed on 2010-03-23)

[5] ISO *ISO Standard 15489 (Information and documentation - Records management) Part 1: General*. ISO, Geneva, 2001.

[6] ISO *ISO Standart 19005-1(Document management - Electronic document file format for long-term preservation) - Part 1: Use of PDF 1.4 (PDF/A)*. ISO, Geneva, 2004.

[7] Kampffmeyer, U. et al *Effiziente Informationsverwaltung mit dem neuen europäischen Records-Management-Standard - MoReq2 und Records Management" (Seminarband)*. PROJECT CONSULT, Hamburg, 2008.

[8] Kampffmeyer, U. *ECM Enterprise Content Management – Whitepaper*. PROJECT CONSULT, Hamburg, 2006.

[9] Körber, N. *Dokumente rüsten für das Archiv* FEiG & PARTNER, Leipzig, 2006. http://www.documanager.de/magazin/artikel_1176_verwaltung_archivierung_digitaler_dokumente.html(Last accessed on 2010-03-06)

[10] Wikipedia *Enterprise Content Management*. Wikipedia, 2004. http://de.wikipedia.org/w/index.php?title=Datei:ECM_Komponenten.jpg (Last accessed on 2010-04-26)