

CITATION

Ward, J.H. (2012). *Managing Data: Preservation Repository Design (the OAIS Reference Model)*. Unpublished Manuscript, University of North Carolina at Chapel Hill.

Creative Commons License: Attribution-NoDerivatives 4.0 International (CC BYND 4.0)

ABSTRACT

In 1995, the Consultative Committee for Space Data Systems (CCSDS) began to coordinate the development of standard terminology and concepts for the long-term archival storage of various types of data. Under the auspices of the CCSDS, experts and stakeholders from academia, government, and research contributed their knowledge to the development of what is now called the Open Archival Information Systems (OAIS) Reference Model. The conclusion from a variety of experienced repository managers is that the authors of the OAIS Reference Model created flexible concepts and common terminology that any repository administrator or manager may use and apply, regardless of content, size, or domain. This literature review summarizes the standard attributes of a preservation repository using the OAIS Reference Model, including criticisms of the current version.

TABLE OF CONTENTS

ABSTRACT	2
INTRODUCTION	5
THE OAIS REFERENCE MODEL: DEFINITION	6
THE OAIS REFERENCE MODEL: KEY CONCEPTS	7
THE OAIS REFERENCE MODEL: KEY RESPONSIBILITIES	12
THE OAIS REFERENCE MODEL: KEY MODELS	12
The OAIS Functional Model	13
The OAIS Information Model	16
I. The Logical Model for Archival Information	17
II. The Logical Model of Information in an Open Archival Information System	18
III. Data Management Information	20
Information Package Transformations	20
THE OAIS REFERENCE MODEL: PRESERVATION PERSPECTIVES	23
Information Preservation	23
Access Service Preservation	24
THE OAIS REFERENCE MODEL: ARCHIVE INTEROPERABILITY	25
THE OAIS REFERENCE MODEL: COMPLIANCE	26
THE OAIS REFERENCE MODEL: EXAMPLE DEPLOYMENTS	28
THE OAIS REFERENCE MODEL: OTHER CRITICISMS	30
CONCLUSIONS AND FUTURE WORK	31
REFERENCES	33

TABLE OF FIGURES

Figure 1 - Environment Model of an OAIS (CCSDS, 2002)	8
Figure 2 - Obtaining Information from Data (CCSDS, 2002)	8
Figure 3 - Information Package Concepts and Relationships (CCSDS, 2002).....	9
Figure 4 - OAIS Archive External Data (CCSDS, 2002).....	11
Figure 5 - OAIS Functional Entities (CCSDS, 2002).....	13
Figure 6 - Composite of Functional Entities (CCSDS, 2002).....	16
Figure 7 - High-Level Data Flows in an OAIS (CCSDS, 2002).....	22

INTRODUCTION

Various organizations and the individuals who work for those organizations have a vested interest in keeping information accessible over time, although there may be reasons to delete or destroy some data and information once a certain amount of time has passed. The reasons for this interest are varied. Librarians and archivists have a professional expectation that they will do their best to curate and preserve cultural heritage data, scientific data, and other types of information for future generations of scholars and laymen. Some interest may be personal -- most people would like to be able to view their children's baby pictures, and their descendants may wish to know how their ancestors looked.

Regardless of the motivation for keeping this information available over time, most practitioners and laymen will agree that standards are one way to ensure this happens. Standards provide a common terminology that aid in discussions of repository infrastructure and needs (Beedham, et al., 2005; Lee, 2010). According to the members of the Science and Technology Council of the Academy of Motion Picture Arts and Sciences (2007), when preservationists and curators collaborate among and between industries and domains to create and use standards, the resulting economy of scale should reduce costs for all involved. For example, Galloway (2004) wrote that the proliferation of file formats increased costs, and that this problem must be solved in order to reduce preservation costs.

If costs are reduced, then the likelihood of a community having the resources to preserve and curate the material increases, or, by the same token, the amount of information that can be saved for the same price increases. This is true across the board, as standards beget other standards. If practitioners and researchers develop a standard terminology for a preservation repository, then common standards for metadata, file formats, filenames, metadata, metadata registries, and archiving and

distributing are likely either to follow or to have preceded the preservation repository standard. In other words, standards development is an iterative process.

In 1995, the Consultative Committee for Space Data Systems (CCSDS) convened to coordinate "the development of archive standards for the long-term storage of archival data" (Beedham, et al., 2005). As part of this task, the members of the CCSDS determined that there was no common model or foundation from which to build an archive standard. Lavoie (2004) describes how the members realized they would have to create terminology and concepts for preservation; characterizations of the functions of a digital archiving system; and determine the attributes of the digital objects to be preserved. Thus, the members agreed to create a reference model that would describe the minimum requirements of an archival system, including terminology, concepts, and system components. The members of the CCSDS recognized from the beginning that the application of a common model extended beyond the space data system, and they involved practitioners and researchers from across a broad spectrum in academia, private industry, and government (Lavoie, 2004; Lee, 2010).

This essay summarizes the standard attributes of a preservation repository as defined by the CCSDS with the Open Archival Information Systems (OAIS) Reference Model, and addresses some of the weaknesses of the model.

THE OAIS REFERENCE MODEL: DEFINITION

An Open Archival Information System (OAIS) is an electronic archive that is maintained by a group or association of people and/or organizations as a system. This member organization has accepted the responsibility of providing access to information for the stakeholders of the electronic archive. These stakeholders are referred to as the Designated Community. The owners and maintainers of the electronic archive have either implicitly or explicitly agreed to preserve the information in the electronic archive

and make it available to the Designated Community for the indefinite long-term (CCSDS, 2002).

The CCSDS created the document for the OAIS Reference Model to outline the responsibilities of the owners and maintainers of the electronic archive. If they meet those responsibilities, then the electronic archive may be referred to as an "OAIS archive". When the CCSDS members used the word "Open" as part of the name of the Reference Model, they referred to the fact that the standard was developed and continues to be developed in open forums. They are clear that the use of the word, "open" does not mean that access to the OAIS system itself or its contents is unrestricted (CCSDS, 2002).

THE OAIS REFERENCE MODEL: KEY CONCEPTS

The members of the CCSDS created three OAIS concepts. They called these the "OAIS Environment", the "OAIS Information", and the "OAIS High-level External Interactions".

The "OAIS Environment" consists of the "Producers", "Consumers", and "Management" in the environment that surrounds an OAIS archive. The "Producer" is a system or people who provide the information (data) that is ingested into the archive to be preserved. The "Consumer" is a system or people who use the archive to access the preserved information. "Management" is a role played by people who are not involved in the day-to-day functioning of the archive, but who also set overall OAIS policy. Other OAIS or non-OAIS compliant archives may interact with the OAIS archive as either a "Producer" or a "Consumer" (CCSDS, 2002). The CCSDS represented these concepts with in Figure 1, below.

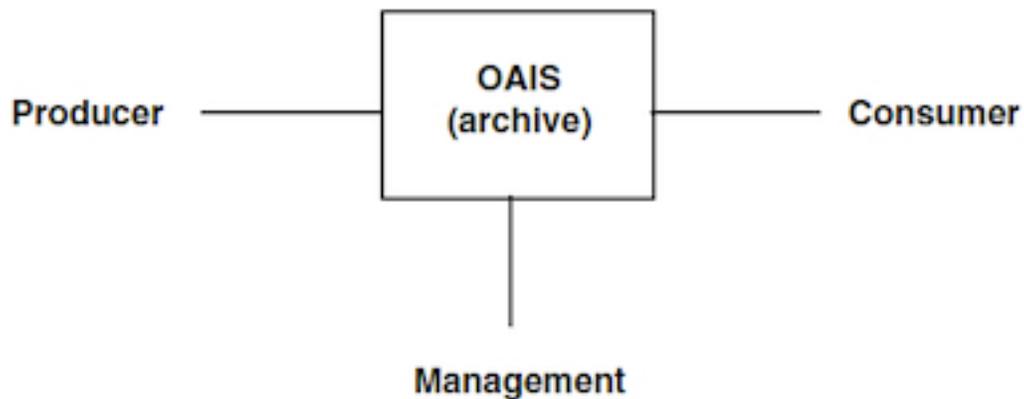


Figure 1 - Environment Model of an OAIS (CCSDS, 2002).

The CCSDS wrote the "OAIS Information" concept to consist of the "information definition", the "information package definition", and the "information package variants".

First, the CCSDS defined "information". Information is "any type of knowledge that can be exchanged, and this information is always expressed (i.e., represented) by some kind of data" (CCSDS, 2002). A person or system's Knowledge Base allows them to understand the received information (see Figure 2, below). Thus, "data interpreted using its Representation Information yields Information" would mean in practice that ASCII characters (the data) representing a language (such as English or French grammar and language, i.e., "the Knowledge Base" or Representation Information) provided Information to the person. Therefore, in order for Information to be represented with any meaning to a Designated Community, the appropriate Representation Information for a Data Object must also be preserved.

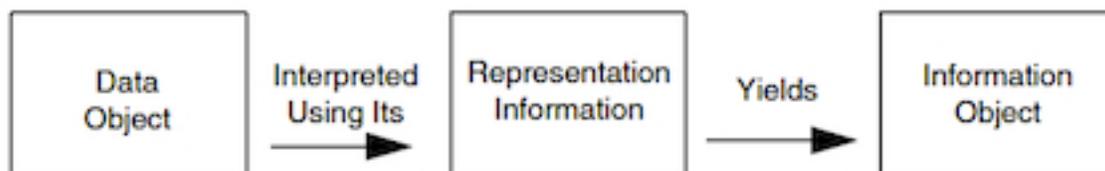


Figure 2 - Obtaining Information from Data (CCSDS, 2002).

Second, whether data is disseminated to a Designated Community member, or ingested via a Producer, the information must be packaged. The CCSDS described an Information Package as consisting of the Packaging information, the Content Information (the information to be preserved and its representation information), and the Preservation Description Information (provenance, context, reference, and fixity). Provenance describes the source of the information; context provides any related information about the object; reference is the unique identifier or set of identifiers for the content; and fixity assures that the content has not been altered, either intentionally or unintentionally. The Packaging Information binds the Content Information and Preservation Description Information, per Figure 3, below.

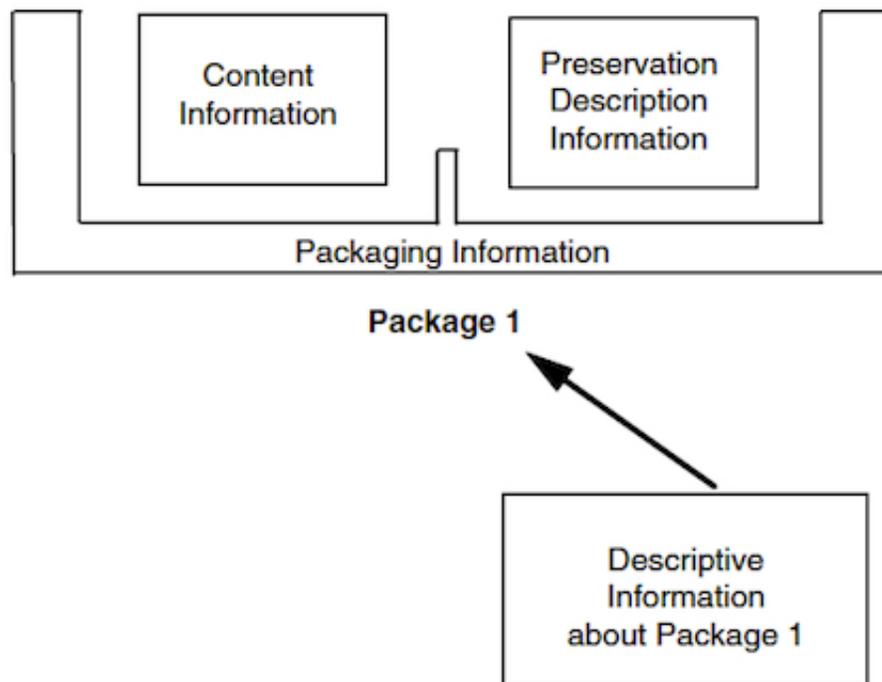


Figure 3 - Information Package Concepts and Relationships (CCSDS, 2002).

Third, the CCSDS defined three variants of the Information Package: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP). These three versions may be the same, but they may also be different. For example, a Producer may submit a SIP to an OAIS archive that is then augmented by the archive managers to meet their policies and standards. Once ingested, the AIP the repository owner stores may or may not be the same as the DIP accessed by the Consumer. Beedham, et al. (2005) criticize the developers of the OAIS Reference Model for assuming that all OAIS archives will have three different versions of an Information Package. The authors note that this concept is not practical for data archives, for example, because all relevant information about a data set must be gathered at the time of submission, and it is impractical to store different versions of an information object within an archive. Thus, a consumer may receive a DIP that is an exact copy of the AIP and the original SIP.

Finally, the CCSDS documented the concepts of the "OAIS High-level External Interactions", in Figure 4, below. In short, they described the external data flows between and among the actors in an "OAIS Environment": management, producer, and consumer. The CCSDS provided example interactions for Management, such as: funding, reviews, pricing policies, and "conflict resolution involving Producers, Consumers, and OAIS internal administration" (CCSDS, 2002).

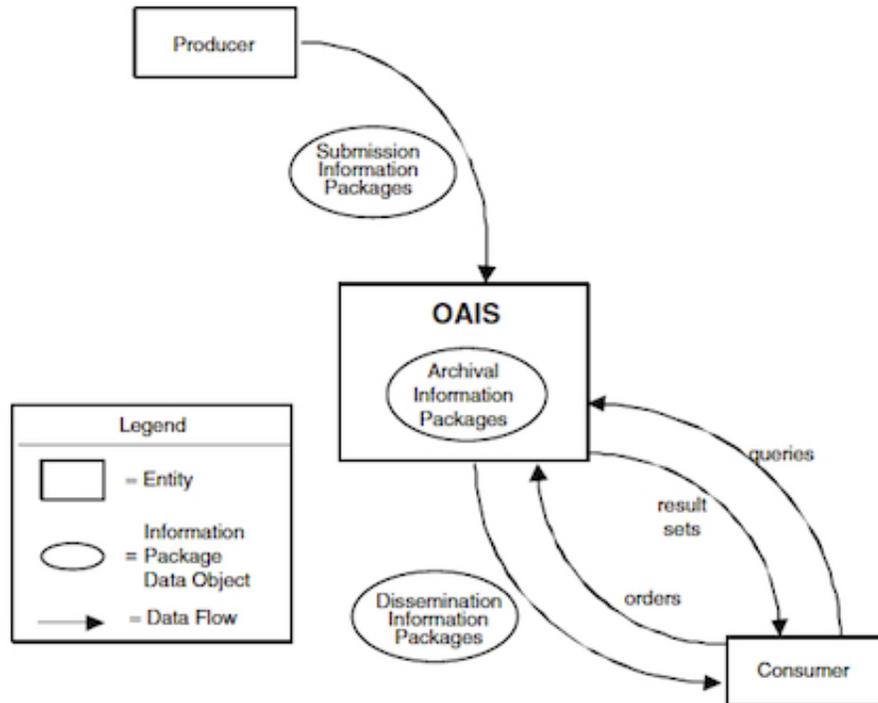


Figure 4 - OAIS Archive External Data (CCSDS, 2002).

The members of the CCSDS described "Producer Interaction" as involving the initial contact, the establishment of a Submission Agreement (which lays out what is to be submitted, how, and other expectations per the two parties) and the Data Submission Session(s) (in which the SIPS are submitted to the OAIS). The authors of the Reference Model conceded that there might be many types of Consumer Interactions with the OAIS managers. They described a variety of interactions, which include catalog searches, orders, help, etc. Beedham, et al. (2005) again criticized the CCSDS for assuming that all OAIS archives will provide order functions to their Designated Communities. The authors point out that some repository's owner policies require that data is available for free, particularly when the owner of the archive is a national government agency, and the Designated Community are taxpayers.

THE OAIS REFERENCE MODEL: KEY RESPONSIBILITIES

The CCSDS established the minimal responsibilities required for a repository to be considered an OAIS archive. The OAIS must:

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided.
- Ensure that the information to be preserved is Independently Understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
- Make the preserved information available to the Designated Community (CCSDS, 2002).

Beedham, et al. (2005) wrote that the authors of the OAIS created an "inbuilt limitation" because they assume "both an identifiable and relatively homogeneous consumer (user) community". They note that this is not the case for national archives and libraries; their Consumers hold a wide variety of skills, educational levels, and knowledge.

THE OAIS REFERENCE MODEL: KEY MODELS

The members of the CCSDS described the functional entities of the OAIS as three models, a "Functional Model", the "Information Model", and "Information Package Transformations". The authors of the Reference Model included this section to provide a common set of preservation system terminology, and to provide a model from which future systems designers may work.

The OAIS Functional Model

The functional model of the OAIS consists of "six functional entities and related interfaces" (CCSDS). The six functional entities are ingest, archival storage, data management, administration, preservation planning, access, and common services. The seventh entity, "Common Services", is described in the document, but it is not included in the image of the OAIS Functional Entities (see Figure 5, below) because "it is so pervasive".

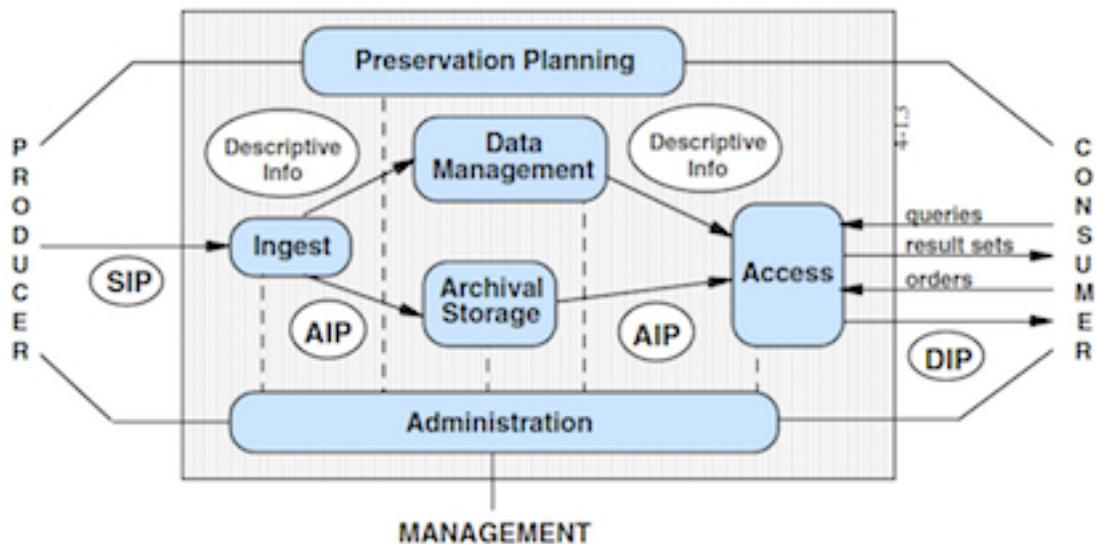


Figure 5 - OAIS Functional Entities (CCSDS, 2002).

1. **INGEST:** Functions of the Ingest entity include accepting SIPs from internal or external Producers and then preparing the SIP(s) for management and storage within the repository. As part of preparing the SIP for storage within the repository, the repository employee in charge of ingest will check the quality of the SIP(s), create an AIP that complies with the standards of the repository and with the Submission Agreement, extract any Descriptive Information, and sync updates between Ingest and Archival Storage/Data Management.

Practitioners such as Beedham, et al. (2005) criticized the lack of detail available for the Ingest process; the authors of the Reference Model made it appear to be a very simple function, when, in fact, it can be a very complex process. As a result of this

criticism, the CCSDS wrote a more detailed description of the Ingest Process in the Producer-Archive Interface Methodology Abstract Standard (CCSDS, 2004). However, many practitioners are clear that "pre-ingest functions are...essential for efficient and effective archiving" and the authors of the OAIS would serve the preservation repository community better by expanding the Ingest section of the OAIS Reference Model documentation, rather than creating a separate model and documentation (Beedham, et al., 2005).

Partially due to the lack of detail related to Ingest, much less the Ingest of records, archivists and records managers at Tufts University and Yale University applied the OAIS Reference Model and developed an Ingest Guide to aid practitioners in preserving university records (Fedora and the Preservation of University Records Project, 2006). (This project was discussed in a previous literature review on digital curation and preservation.)

2. **ARCHIVAL STORAGE:** Functions of archival storage include maintaining the integrity of the digital files, including the bits. Thus, the functions of this entity include not only receiving the AIP from Ingest and to Access, but also refreshing and migrating the media and file formats on and in which the data is stored. Other tasks of this entity include error checking and disaster recovery.
3. **DATA MANAGEMENT:** The data management entity provides the functions and services for accessing, maintaining, and populating administrative data and Descriptive Information. These include generating reports from result sets which are based on queries on the data management data; updating the database; and maintaining and administering archive database functions, such as referential integrity and view/schema definitions.

Beedham, et al. (2005) concluded that this entity is a simple idea that is messy in practice. When they mapped the different data management entities, their results created an "explosion" to all the different archival systems and processes.

4. **ADMINISTRATION:** The functions of this entity involve the overall management of the archive. This includes setting policies and standards; supporting and aiding the Designated Community; migrating and refreshing the archive contents,

software, and hardware; and soliciting, negotiating, auditing Submission Agreements with both internal and external producers; and, any other administrative related duties as required.

These functions are designed for large organizations with automated processes; the authors of the Reference Model did not design this entity for small-scale digital repositories (Beedham, et al., 2005). However, most of these functions are an organic part of many archive's functioning, even if the roles are all performed by one or two people. Beedham, et al. (2005) wrote that the functions of this entity are sufficient for most archives, but the listed tasks do not stand on their own, as each archive has its own set of responsibilities, requirements, procedures, and policies.

- 5. PRESERVATION PLANNING:** This preservation planning entity is related to the Administrative entity, but it focuses purely on the preservation aspects of maintaining the archive for the indefinite long-term and ensuring the content is available to the Designated Community. The functions of the entity primarily involve monitoring the internal and external environments of the archive to ensure hardware and software are up to date; that the archive follows best practices with regards to the preservation of digital content; and that plans are in place to enable Administration goals, such as migration.

Repository managers criticized this entity because "real" archives do not operate as cleanly as the OAIS Reference Model authors envision; not all decisions and processes can or should be made proactively. Beedham, et al. (2005) concluded that the OAIS is at times overly bureaucratic and formalized.

- 6. ACCESS:** This function provides the Designated Community with a method to obtain the desired information from the archive, assuming such access is not restricted and that the user in question is, in fact, allowed to access this particular information from this particular archive. The services and functions provided by the Access entity allow the Designated Community to determine the existence, location, availability, and description of the stored information. This function provides the information to the Designated Community as a DIP.
- 7. COMMON SERVICES:** The "common services" functional entity refers to supporting services common in a distributed computing environment. These services involve operating systems, network services, and security services. Operating system services include the core services required to administer and operate an application platform, and provide an interface. These include: system management, operating system security services, real-time extension,

commands and utilities, and, kernel operations. Network services provide the means for the archive to operate in a distributed network environment, including: remote procedure calls, network security services, interoperability with other systems, file access, and data communication. Security services protect the content in the archive from external and internal threats by providing the following capabilities and mechanisms: non-repudiation services (i.e., the sender and receiver log copies of the transmission and receipt of the information), data confidentiality and integrity services, access control services, and authentication (CCSDS, 2002).

A detailed mapping of the Ingest, Archival Storage, Data Management, Administration, Preservation Planning, and Access functional entities is included in Figure 6, below.

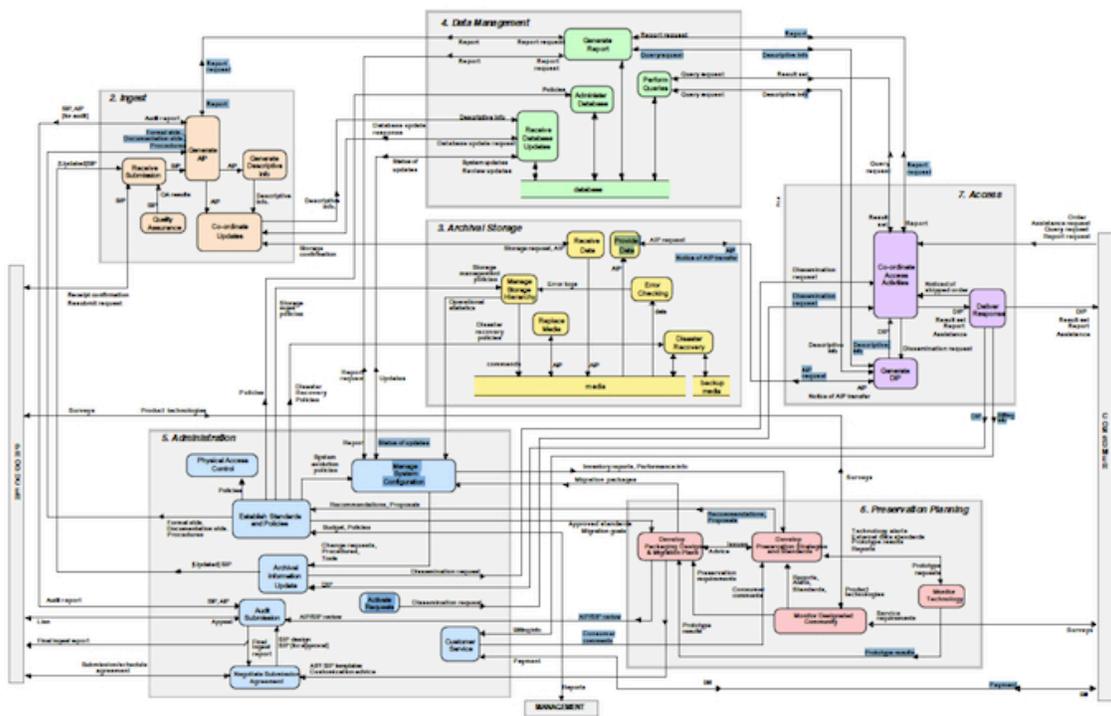


Figure 6 - Composite of Functional Entities (CCSDS, 2002).

Again, Common Services is not included because it is a supporting service of distributed computing (CCSDS, 2002).

The OAIS Information Model

The Information Model "defines the specific Information Objects that are used within the OAIS to preserve and access the information entrusted to the archive"

(CCSDS, 2002). The CCSDS intended for this section to be conceptual, and it is written for an Information Architect to use when designing an OAIS-compliant system. The authors divided the Information Model into three sections: the logical model for archival information, the logical model of information in an open archival information system (OAIS), and data management information.

1. The Logical Model for Archival Information

The CCSDS defined information as a combination of data and representation information. The Information Object itself is either a physical or digital Data Object with Representation Information that "allows for the full interpretation of data into meaningful information" (CCSDS, 2002). The Representation Information provides a method for the data to be mapped to data types such as pixels, arrays, tables, numbers, and characters. The latter are referred to as the Structure Information and Semantic Information, in turn, supplements this. Semantic Information examples include the language expressed in the Structure Information, which kinds of operations may be performed on each data type, their interrelationships, etc. Representation Information may also reference other Representation Information; for example, "Representation Information expressed in ASCII needs the additional Representation Information for ASCII, which might be a physical document giving the ASCII Standard" (CCSDS, 2002).

Representation Rendering Software and Access software are two special types of Representation Information. The latter provides a method for some or all of the content of an Information Object to be in a form understandable to systems or a human. The former displays the Representation Information in an understandable form, such as a file and directory structure (CCSDS, 2002).

The CCSDS defined four types of Information Objects: Content, Preservation Description, Packaging, and Descriptive. The Content Information Object is "the set of information that is the original target of preservation by the OAIS" and it may be either a

physical or digital object (CCSDS, 2002). In order to determine clearly what must be preserved, an administrator of an archive must determine which part of a Content Information Object is the Content Data Object and which part is the Representation Information.

The CCSDS defined Preservation Descriptive Information as "information that will allow the understanding of the Content Information over an indefinite period of time" (CCSDS, 2002). This descriptive information focuses on ensuring the authenticity and provenance of the Information Objects. The authors of the Reference Model described four parts to the Preservation Descriptive Information: reference (unique identifier(s)), context (why it was created and how it relates to other Information Objects), provenance (the history, origin, and source), and fixity (data integrity checks or validation/verification keys).

As stated previously, the Packaging Information logically binds the pieces of the package onto a specific media via an identifiable entity. Finally, Descriptive Information provides a method for the Designated Community to locate, analyze, retrieve, or order the desired information via some type of Access Aid, which is generally an application interface or document (CCSDS, 2002).

II. The Logical Model of Information in an Open Archival Information System

The authors of the Reference Model described three types of Information Packages that are based on the four types of Information Objects. That is, the Content, Preservation Description, Packaging, and Descriptive Information Objects may be used to create one of three types of Information Packages: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP). The SIP is the data that is sent to an archive by an internal or external Producer. The form and content of the SIP(s) may or may not meet the requirements of the archive ingesting it, and the archive manager may require some

additional information to be added prior to ingest, such as a unique ID, checksum validation, virus checks, file name standardization, or additional Representation Information (metadata).

The CCSDS defined the AIP as the Information Package that is stored for the indefinite long-term. The requirements for the Representation Information for an AIP are more stringent than for other types of Information Packages, because this is the actual information that is the focus of preservation. The Information Objects and the Representation Information that comprise an AIP are stored in an archive as one logical unit (Lavoie, 2004).

The authors described two subsets of the AIP, the Archival Information Unit (AIU) and the Archive Information Collection (AIC). The former "represents the type used for the preservation function of a single content atomic object", while the latter "organizes a set of AIPs (AIUs and other AICs) along a thematic hierarchy..." (CCSDS, 2002). The CCSDS described the Collection Description as a subtype..."that has added structures to better handle the complex content information of an AIC" (CCSDS, 2002). The archive manager may use Collection Description to describe the entire collection or zero or more individual units within the collection. One benefit of Collection Description is the ability to generate new virtual collections based, for example, either on access or theme.

The Dissemination Information Package (DIP) is the Information Package ordered by or provided to the Designated Community. The CCSDS intended for the DIP to be a version of the AIP, but it is entirely possible for the AIP and the DIP to be exactly the same Information Package. Lavoie (2004) described possible variations between an AIP and a DIP. The Designated Community member accessing the archive may receive a different format, for example, a .jpeg instead of a .tiff. The DIP may contain less metadata than is available with the AIP, or even less content, since a DIP may correspond to one or more or even part of an AIP.

III. Data Management Information

Last, the CCSDS included Data Management Information as one part of the Logical Model of Information in an OAIS. That is, the authors of the Reference Model made the requirement that information needed for the operation of the archive is to be stored in the archive databases as persistent data classes. The type of information required includes: statistical information, such as access numbers; customer profile information; accounting information; preservation process history; event based order information; policy information, including pricing; security information; and, transaction tracking information (CCSDS, 2002). Other data management information may be added to the archive at the discretion of the archive managers or as requested by the Designated Community. However, Beedham, et al. (2005) concluded that the information categories in the Information Model are "too broad, functionally organised...and do not reflect the way metadata are packaged and used across particular archival practice".

Information Package Transformations

The CCSDS members created the Functional Model to describe the architecture of an OAIS, and the Information Model to describe the content held by the OAIS. The authors also described the lifecycle of the Information Package and any associated objects, as well as its logical and physical transformations.

In short, when a Producer agrees to submit data to an OAIS, a Submission Agreement is created and approved with the OAIS administrator. The Producer then submits data in the form of a SIP to an OAIS, where the OAIS administrator stores it in a staging area. In the staging area, the OAIS manager will perform any necessary transformations to the SIP so it will meet the standards of the OAIS, and the criteria of the Submission Agreement. The OAIS manager will create AIPs from the SIP. This mapping may not be one-to-one. One SIP may produce one AIP or many AIPs, many

SIPs may produce one AIP, many SIPs may produce many AIPs, and one SIP may produce *no* AIPs (CCSDS, 2002). The CCSDS described this process in more detail in the *Producer-Archive Interface Methodology Abstract Standard* (CCSDS, 2004).

At the same time as the SIPs are transformed into AIPs and stored in the OAIS, the Data Management functional entity augments the existing Collection Descriptions to include the contents of the Package Descriptions. When a Consumer, i.e., a member of the Designated Community, wishes to access the information contained in an OAIS, the member will do so via the Access functional area. Once the consumer has located the desired information via some type of finding aid, the information is provided to the Consumer in the form of a DIP. The authors of the Reference Model designed the DIP and AIP mapping to be similar to that between SIPs and AIPs. That is, the mapping may or may not be 1:1, depending on whether or not a transformation is performed.

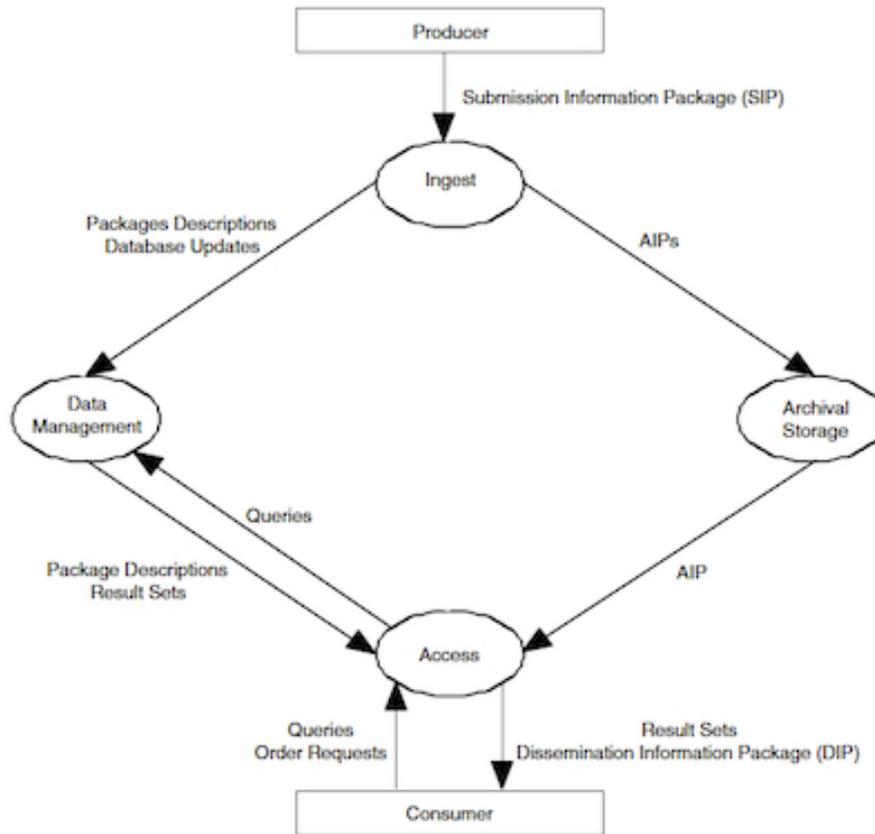


Figure 7 - High-Level Data Flows in an OAIS (CCSDS, 2002).

Based on the Information Package Transformation in Figure 7, above, the authors of the OAIS Reference Model assumed that the Consumer from an AIP would create a DIP on demand. Beedham, et al. (2005) wrote, "this approach has serious drawbacks". These data repository managers determined that by creating the DIP at the time of Ingest, they could ensure that the records accessed by the Consumer are in a "technically usable state" (Beedham, et al., 2005). They initially created DIPS from an AIP upon demand by a Consumer, but often times, the data is 5-10 years old at the time of ingest into the archive, and the data is often years older than that when accessed by a Consumer. This often meant that the DIP was not independently understandable by the Consumer, and the researchers who created the data either were no longer available, or could not answer queries regarding the data because too much time had passed.

Beedham, et al., (2005) discovered that by creating the DIP at the time of Ingest, they were able to eliminate many errors in the digital records while they still had co-operation from the Producer. This also improves the understanding and “preservability” of the AIP itself. As well, standard archival practice is to store the original version, and provide only a copy to users. In that sense, storing the AIP and creating a DIP at Ingest that is an exact replica of the AIP follows this practice, although "copy" does not have the same meaning in the digital world as it does in the physical. The OAIS Reference Model does not preclude this practice, but neither does it explicitly condone it.

THE OAIS REFERENCE MODEL: PRESERVATION PERSPECTIVES

The members of the CCSDS used the Functional Model and the Information Model just described and applied them to information preservation and access service preservation. The former refers to the migration of digital information and the latter to the preservation of the services used to access the digital information.

Information Preservation

The CCSDS defined migration as "the transfer of digital information, while intending to preserve it, within the OAIS" (CCSDS, 2002). The authors distinguished migration from transfers based on three characteristics: the focus is on the preservation of the full information content; the new archival implementation is a replacement for the old; the responsibility for and full control of the transfer reside within the OAIS. The CCSDS (2002) members described three primary drivers for migration: the media on which the information resides is decaying; technology changes; and, the improved cost-effectiveness of newer technology over older or obsolete technology.

The committee members defined four types of migration: refreshment, replication, repackaging, and transformation. They determined that Refreshment refers to the replacement of a media instance with a similar piece of media, such that the bits

comprising the AIP are simply copied over. An example of this would be replacing a computer disk. The authors defined Replication as a bit transfer to the same or new media-type, where there is "no change to the PDI, the Content Information, and the Packaging Information. An example of replication would be a full back up of the contents of an OAIS. The CCSDS described Repackaging as a change to the Packaging Information during transfer. If files from a CD-ROM are moved to new files on another media type, with a new file implementation and directory, then the files have been Repackaged.

Last, the CCSDS (2002) defined Transformation as "some change in the Content Information or PDI bits while attempting to preserve the full information content". If an AIP undergoes Transformation, then the new AIP is considered a new Version of the previous AIP. For example, a file in the .doc format may be transformed to a .pdf for preservation purposes. Some transformations are Reversible, while others are Non-reversible. The CCSDS members state that only when an AIP is migrated using Transformation is the resulting AIP considered a new version; the AIP version is independent of Refreshment, Replication, and Repackaging.

Access Service Preservation

As part of examining preservation perspectives, the members of the CCSDS briefly addressed how to continue to provide Consumers access services as technology changes. A method archive managers use to maintain access is to develop Application Programming Interfaces (APIs) to provide access to AIPs. Another method they incorporate is to use emulation or provide the original source code to provide access to a set of AIUs while maintaining the same "look and feel" as the original access method.

THE OAIS REFERENCE MODEL: ARCHIVE INTEROPERABILITY

A community of users and managers of digital repositories may wish to share data or cooperate with other archives. The reasons for this may vary; in some cases, the repository managers may wish to provide mutual back up and replication services with a similar archive, in order to prevent data loss and reduce costs. In another instance, a user community may prefer one point of entry to search for required information across multiple digital archives. Regardless of the motivations of an archive owner for interoperating with another archive, the interactions may be defined by two categories, technical and managerial.

The CCSDS defined four types of interoperating archives: independent, cooperating, federated, and shared resources. They described an independent archive as one that does not interact with other archives. There is no technical or management interaction between this type of archive and other archives. The authors defined cooperating archives as those archives that do not have a common finding aid, but otherwise share common dissemination standards, submission standards, and producers.

The members of CCSDS (2002) wrote that a federated archive consists of two communities, Local and Global, and those archives "provide access to their holdings via one or more common finding aids". They note that Global dissemination and Ingest are optional, and that the needs of the Local community tend to take precedence over the Global community. Furthermore, they described three levels of functionality for a Federated archive: Central Site (i.e., one point of entry to all archive content via metadata harvested by the central site), Distributed Finding Aid (i.e., federated searching of all archives), and Distributed Access Aid (i.e., a "standard ordering and dissemination mechanism") (CCSDS, 2002). They wrote that federated archives tend to have similar

policy and technology issues, such as authentication and access management, preservation of federation access to AIPs, duplicate AIPs, and providing unique AIPs.

Last, the authors described "shared resources", where archives enter into agreements to share resources for their mutual benefit, often to reduce costs. They wrote that this type of agreement does not alter the view of the archives by their respective Designated Communities, it merely requires the implementation of a variety of standards internal to the archive, such as ingest-storage and access-storage interface standards (CCSDS, 2002).

The CCSDS described the primary management issue related to archive interoperability in one word: autonomy. The members of the CCSDS (2002) characterized three primary autonomy levels: no association because there are no interactions; an association member's autonomy with regards to the federation is maintained; and association members are bound to the federation by a contract.

THE OAIS REFERENCE MODEL: COMPLIANCE

What does it mean to be "OAIS Compliant"? The members of the CCSDS stated that if a repository "supports the OAIS information model", commits to "fulfilling the responsibilities listed in chapter 3.1 of the reference model", and uses the OAIS terminology and concept appropriately, then the archive is compliant (CCSDS, 2002; Beedham, et al., 2005). When the members of the CCSDS wrote the Reference Model documentation, they did not recommend any particular concrete implementation of hardware, software, etc., as the authors deliberately designed it to be a conceptual framework. How then, may an archive owner, manager, or member of a Designated Community "prove" that the archive of interest is, in fact, OAIS-compliant?

One method to audit OAIS-compliance is to create a set of standards that define the attributes of a trusted digital repository. The Research Libraries Group (RLG) and the

Online Computer Library Center (OCLC) funded the development of the attributes of a "trusted digital repository" in March 2000. The two groups produced a report that defined the attributes and responsibilities of a trusted digital repository in 2002 (Research Libraries Group, 2002). Beedham, et al. (2005) notes that the authors of the report put compliance with the OAIS Reference Model first on the list of attributes of a trustworthy repository.

Based on this report, RLG, OCLC, the Center for Research Libraries (CRL), and the National Archives and Records Administration (NARA) produced a "criteria and checklist" in 2005 called, "Trustworthy Repositories Audit & Certification: Criteria and Checklist" (Research Libraries Group, 2005). The authors designed it so that archive managers could use it for audit and certification of the archive. Experts in the field merged the RLG and OCLC report from 2002 and the "Criteria and Checklist" from 2005 to develop a Recommended Practice under the auspices of the CCSDS. They called the document the "Audit and Certification of Trustworthy Digital Repositories Recommended Practice", and the CCSDS released the document in September 2011. The CCSDS released the document to provide a basis for the audit and certification of the trustworthiness of a digital repository by providing detailed criteria by which an archive shall be audited (CCSDS, 2011). These documents will be discussed in detail in a separate literature review.

One criticism of the OAIS is that it is challenging to develop a from-scratch repository using the Reference Model. Egger (2006) conducted a use case analysis as part of a standard software development process, and determined that he must "develop additional specifications which fill the gap between the OAIS model and software development". He wrote that it was difficult to map OAIS functions as use case scenarios, because the descriptions contain different levels of detail. For example, he states that some functions are written as general guidelines, while others are "specified

nearly at the implementation level" (Egger, 2006). He also criticizes the authors for mixing technical functionality with management functionality, because in order to develop a technical system, the management functions must be removed. Egger (2006) recommends creating additional specifications that would "define system architectures and designs that conform to the OAIS model", although he notes that the OAIS Reference Model is not a technical guideline.

Beedham, et al. (2005) wrote that as repository managers, they have to consider other legislation, standards, guidelines, and regulations when determining the archive's OAIS compliance. For example, they must provide web access to the disabled as part of their charter as national archives, and they have specific responsibilities to the data depositor (the Producer) with regards to Intellectual Property and statistical disclosure. The authors of the Reference Model did not discuss how to comply with legislation, et al., when to do so would make the archive in question "not OAIS-compliant", if audited.

THE OAIS REFERENCE MODEL: EXAMPLE DEPLOYMENTS

Ball (2006) examined the OAIS Reference Model to determine the application of it to engineering repositories. Two common generic repository systems that use the OAIS Reference Model are DSpace¹ and Fedora². The creators of DSpace designed it primarily for Institutional Repositories, while the researchers behind Fedora designed it to be a digital library that stores multimedia collections. Ball found five custom repositories that claim to be OAIS-compliant: the Centre de Données de la Physique des Plasmas³ (CDPP), MathArc⁴, the European Space Agency (ESA) Multi-Mission Facility Infrastructure⁵ (MMFI), the National Oceanic and Atmospheric Administration (NOAA)

¹ <http://www.dspace.org/>

² <http://fedora-commons.org/>

³ <http://cdpp.cesr.fr/>

⁴ <http://www.library.cornell.edu/dlit/MathArc/web/index.html>

⁵ http://www.dlr.de/caf/en/desktopdefault.aspx/tabid-5387/8999_read-16749/

Comprehensive Large Array-data Stewardship System⁶ (CLASS), and, the National Space Science Data Center⁷ (NSSDC). While Ball did discuss the efforts of RLG, OCLC, CRL, and NARA to provide a method for audit and certification, he did not note whether or not the creators and owners of DSpace, Fedora, or any of the custom systems, or their users, had formally audited any of the repository software for OAIS compliance.

Vardigan & Whiteman (2007) did apply the OAIS Reference Model to the social science data archive for the Inter-university Consortium for Political and Social Research⁸ (ICPSR). The authors wished to determine their repository's conformance to the OAIS Reference Model. After an extensive audit, they realized that the ICPSR digital repository did fulfill many of the key responsibilities of an OAIS archive, with two exceptions. First, they need to publish a preservation policy, and second, they discovered that their Preservation Description Information is not always clearly labeled and it is often incomplete (Vardigan & Whiteman, 2007).

Data grids are an example of a general systems deployment of the OAIS Reference Model. A grid administrator may map the policies and procedures that govern the data flow of the data grid to specific OAIS components. For example, if the grid administrator would like to create authentic copies, then s/he will implement access policies that govern the generation of DIPs. The grid administrator may implement replication and integrity checking by implementing storage policies; and may implement the processing of SIPs and the creation of AIPs by implementing ingest policies (Reagan Moore, personal communication, December 22, 2011). Other specific OAIS components may be mapped to the data grid's policies and procedures data flow as needed; these are but a few examples.

⁶ <http://www.class.noaa.gov/>

⁷ <http://nssdc.gsfc.nasa.gov/>

⁸ <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

THE OAIS REFERENCE MODEL: OTHER CRITICISMS

Higgins and Semple (2006) compiled a list of recommendations for updates to the OAIS Reference Model in preparation for the CCSDS' review of the recommendation at the five-year mark. The authors compiled the list of recommendations on behalf of the Digital Curation Centre and the Digital Preservation Coalition. Among the general recommendations, the authors listed: supplementary documents such as OAIS-lite for managers, a self-testing manual, an implementation checklist, and a best practice guide. The authors requested more concrete and up-to-date examples for implementers.

Higgins and Semple noted the CCSDS' tendency to be very prescriptive and detailed in some sections, and overly general in others. They re-iterated that the CCSDS should create a better description of minimal requirements, as not everything must be implemented. The authors requested a review of the terminology clashes between the OAIS Reference Model, PREMIS, and other standards, and asked the CCSDS to resolve these differences. Higgins and Semple requested terminology and clarification updates by chapter, including updates to words such as "repository", "preservation", "security", etc. They also identified a variety of outdated material.

The members of the CCSDS Data Archiving and Ingest Working Group did respond to this list of recommendations. They adopted some of the recommendations and made changes to the text of the OAIS Reference Model, but they refused to make other requested changes. Higgins and Boyle (2008) compiled a response to the CCSDS, again on behalf of the Digital Curation Centre and the Digital Preservation Coalition. Their concerns related to the changes rejected by the CCSDS Data Archiving and Ingest Working Group. Higgins and Boyle (2008) wanted "to ensure that the revised standard" would:

- remain up-to-date until the next planned review;
- remain applicable to the current heterogeneous user base;

- be easier to understand through a structure which clearly delimits normative text, use cases and examples;
- contain guidelines on how to achieve an implementation;
- follow ISO practice by clearly referencing other applicable standards; and,
- clarify its applicability to digital material (Higgins & Boyle, 2008).

It will be interesting to note which, if any, of these recommendations the members of the CCSDS include in the next revision of the OAIS Reference Model.

CONCLUSIONS AND FUTURE WORK

Practitioners note that one benefit of the OAIS Reference Model has been "the utility of the OAIS language as a means of communication" between partnering repository administrators, who often had different terminology (Beedham, et al., 2005). The authors recommend that current archives should adopt the OAIS language in lieu of their own terminology, and new archive administrators should adopt it from the inception of the archive. Allinson (2006) writes that the OAIS Reference Model "ensures good practice", as it "draws attention to the important role of preservation repositories" by providing a standard model so that preservation is considered part and parcel of other archive functions and activities. When the CCSDS outlined an archive manager's Mandatory Responsibilities, the authors asked only that an archive's "preservation has been planned for and a strategy identified", as most repository managers already fulfill those tasks as a de facto part of the repository's functioning (Allinson, 2006).

One area of future work may be to create an "OAIS lite" for smaller archives, who do not have the personnel or need for such a bureaucratic model (Beedham, et al., 2005). Another area for future work is to de-homogenize the definition of Designated Community, as not every repository has a narrow audience of users. The CCSDS might consider recommending other metadata documentation to supplement the Reference Model, or create a separate recommendation; similar to the way the Producer-Archive Interface Methodology Abstract Standard (CCSDS, 2004) supplements the Ingest entity.

This documentation would describe how the different information packages breakdown or how to apply metadata schemas (Beedham, et al., 2005; Allinson, 2006).

Egger (2006), Allinson (2006), and Beedham, et al. (2005), among others, complained that the authors of the OAIS Reference Model are inconsistent in the specifications, as some specifications are very general, while others are very detailed. Therefore, one area for future work is for the CCSDS to create consistency within the Reference Model document with regards to specificity. Finally, Beedham, et al., concluded that the authors of the Reference Model may want to re-word the recommendation to take into account that a SIP, AIP, and DIP may all be one and the same, rather than assume that each of these are different types of Information Packages.

In spite of the various criticisms, the overall conclusion from a variety of experienced repository managers is that the authors of the OAIS Reference Model created flexible concepts and common terminology that any repository administrator or manager may use and apply, regardless of content, size, or domain (e.g., academia, private industry, and government).

REFERENCES

- Allinson, J. (2006). *OAIS as a reference model for repositories an evaluation*. Bath, England: UKOLN. Retrieved December 19, 2011, from <http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/Drs-OAIS-evaluation-0.5.pdf>
- Ball, A. (2006). *Briefing paper: the OAIS Reference Model*. Bath, England: UKOLN. Retrieved December 19, 2011, from <http://homes.ukoln.ac.uk/~ab318/docs/ball2006oais/>
- Beedham, H., Missen, J., Palmer, M. & Ruusalepp, R. (2005). *Assessment of UKDA and TNA compliance with OAIS and METS standards*. UK Data Archive and The National Archives, 2005. Retrieved: December 20, 2011, from: http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf
- CCSDS. (2002). *Reference model for an Open Archival Information System (OAIS) (CCSDS 650.0-B-1)*. Washington, DC: National Aeronautics and Space Administration (NASA). Retrieved April 3, 2007, from <http://nost.gsfc.nasa.gov/isoas/>
- CCSDS. (2004). *Producer-archive interface methodology abstract standard (CCSDS 651.0-B-1)*. Washington, DC: National Aeronautics and Space Administration (NASA). Retrieved August 18, 2007, from <http://public.ccsds.org/publications/archive/651x0b1.pdf>
- CCSDS. (2011). *Audit and Certification of Trustworthy Digital Repositories (CCSDS 652.0-M-1)*. Magenta Book, September 2011. Washington, DC: National Aeronautics and Space Administration (NASA).
- Egger, A. (2006). Shortcomings of the Reference Model for an Open Archival Information System (OAIS). *IEEE TCDL Bulletin*, 2(2). Retrieved October 23, 2009, from <http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html>
- Fedora and the Preservation of University Records Project. (2006). *2.1 Ingest Guide, Version 1.0* (tufts:central:dca:UA069:UA069.004.001.00006). Retrieved April 16, 2009, from the Tufts University, Digital Collections and Archives, Tufts Digital Library Web site: <http://repository01.lib.tufts.edu:8080/fedora/get/tufts:UA069.004.001.00006/bdef:TuftsPDF/getPDF>
- Galloway, P. (2004). Preservation of digital objects. In B. Cronin (Ed.), *Annual Review of Information Science and Technology*, 38(1), (pp. 549-590).
- Higgins, S. & Boyle, F. (2008). *Responses to CCSDS' comments on the 'OAIS five-year review: recommendations for update 2006'*. London: Digital Curation Center and Digital Preservation Coalition.
- Higgins, S. & Semple, N. (2006). *OAIS five-year review: recommendations for update*. London: Digital Curation Center and Digital Preservation Coalition.
- Lavoie, B. (2004). *The open archival information system reference model: introductory guide*. Technology Watch Report. Dublin, OH: Digital Preservation Coalition. Retrieved March 6, 2007, http://www.dpconline.org/docs/lavoie_OAIS.pdf

Lee, C. (2010). Open archival information system (OAIS) reference model. In *Encyclopedia of Library and Information Sciences*, Third Edition. London: Taylor & Francis.

Research Libraries Group. (2002). *Trusted digital repositories: attributes and responsibilities an RLG-OCLC report*. Mountain View, CA: Research Libraries Group. Retrieved September 11, 2007, from <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>

Research Libraries Group. (2005). *An audit checklist for the certification of trusted digital repositories, draft for public comment*. Mountain View, CA: Research Libraries Group. Retrieved April 14, 2009, from <http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2416.pdf>

Science and Technology Council. (2007). *The digital dilemma strategic issues in archiving and accessing digital motion picture materials*. The Science and Technology Council of the Academy of Motion Picture Arts and Sciences. Hollywood, CA: Academy of Motion Picture Arts and Sciences.

Vardigan, M. & Whiteman, C. (2007). *ICPSR meets OAIS: applying the OAIS reference model to the social science archive context*. *Archival Science*, 7(1). Netherlands: Springer. Retrieved February 20, 2008, from <http://www.springerlink.com/content/50746212r6g21326/>