

The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)

Brian Lavoie
OCLC Research

DPC Technology Watch Report 14-02 October 2014

Series editors on behalf of the DPC
Charles Beagrie Ltd.



Principal Investigator for the Series
Neil Beagrie



© Digital Preservation Coalition 2014 and Brian Lavoie 2014

ISSN: 2048-7916

DOI: <http://dx.doi.org/10.7207/twr14-02>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing from the publisher. The moral rights of the authors have been asserted.

First published in Great Britain in 2004

Second edition 2014, the Digital Preservation Coalition.

Foreword

The Digital Preservation Coalition (DPC) is an advocate and catalyst for digital preservation, ensuring our members can deliver resilient long-term access to digital content and services. It is a not-for-profit membership organization whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. It supports its members through knowledge exchange, capacity building, assurance, advocacy and partnership. The DPC's vision is to make our digital memory accessible tomorrow.

The *DPC Technology Watch Reports* identify, delineate, monitor and address topics that have a major bearing on ensuring our collected digital memory will be available tomorrow. They provide an advanced introduction in order to support those charged with ensuring a robust digital memory, and they are of general interest to a wide and international audience with interests in computing, information management, collections management and technology. The reports are commissioned after consultation among DPC members about shared priorities and challenges; they are commissioned from experts; and they are thoroughly scrutinized by peers before being released. The authors are asked to provide reports that are informed, current, concise and balanced; that lower the barriers to participation in digital preservation; and that are of wide utility. The reports are a distinctive and lasting contribution to the dissemination of good practice in digital preservation.

This report was written by Brian Lavoie, Research Scientist at OCLC. The report is published by the DPC in association with Charles Beagrie Ltd. Neil Beagrie, Director of Consultancy at Charles Beagrie Ltd, was commissioned to act as principal investigator for, and managing editor of, this Series in 2011. He has been further supported by an Editorial Board drawn from DPC members and peer reviewers who comment on text prior to release: William Kilbride (Chair), Janet Delve (University of Portsmouth), Marc Fresko (Inforesight), Sarah Higgins (University of Aberystwyth), Tim Keefe (Trinity College Dublin), and Dave Thompson (Wellcome Library).

Acknowledgements

The author would like to thank Neil Beagrie, editor of the *DPC Technology Watch Reports* series, for his patient encouragement in steering this writing project to completion, and also several anonymous reviewers for their helpful commentary and suggestions on an earlier draft of this report.

Contents

1.	Abstract.....	1
2.	Executive Summary.....	2
3.	Introduction	4
4.	Development of the OAIS Reference Model.....	5
5.	Overview of the OAIS Reference Model	7
5.1.	Open Archival Information System.....	7
5.2.	OAIS Environment	9
5.3.	OAIS Functional Model	11
5.4.	OAIS Information Model.....	14
5.5.	Archive Interoperability.....	19
6.	Impact of the OAIS Reference Model	21
6.1.	OAIS-compliant repository architectures	21
6.2.	Repository certification	23
6.3.	Metadata for digital preservation	24
6.4.	Encoding and exchanging archived information	26
6.5.	Other OAIS-related standards	27
7.	Conclusions	29
7.1.	The OAIS legacy so far	29
7.2.	Some limitations.....	30
7.3.	OAIS: a theory of digital preservation	31
8.	Glossary of Acronyms	32
9.	References	33

1. Abstract

Originally developed as part of a broader effort to develop formal standards for the long-term storage of digital data generated from space missions, the Open Archival Information System (OAIS) has since formed the foundation of numerous architectures, standards, and protocols, influencing system design, metadata requirements, certification, and other issues central to digital preservation. This *Technology Watch Report* traces the history, salient features, and impact of the OAIS reference model. It is suitable both as a gentle introduction to OAIS for those new to the reference model, or as a resource for practitioners wishing to reacquaint themselves with the basics of the model and subsequent developments. The report concludes with a brief discussion of OAIS's key benefits and limitations, drawn from the model's legacy of more than a decade of use in the digital preservation community.

2. Executive Summary

The Consultative Committee for Space Data Systems (CCSDS) initiated work aimed at developing formal standards for the long-term storage of digital data generated from space missions. Part of this effort involved the development of a reference model for an 'open archival information system' (OAIS). The reference model would represent a comprehensive and consistent framework for describing and analysing digital preservation issues, provide a sound footing for future standards-building activity, and serve as a point of reference for vendors interested in building digital preservation products and services. The OAIS reference model was approved in January 2002 as ISO International Standard 14721; a revised and updated version was published in 2012 as ISO Standard 14721:2012.

The central concept in the reference model is that of an open archival information system. An OAIS-type archive must meet a set of six minimum responsibilities to do with the ingest, preservation, and dissemination of archived materials.

An OAIS-type archive operates in an environment populated by three types of entities: Management, Producer, and Consumer. A special class of Consumer is called the Designated Community: the subset of Consumers expected to independently understand the archived information in the form in which it is preserved and made available by the OAIS. An OAIS-type archive's external environment could also include interaction with other OAIS archives.

The reference model identifies and describes the core set of mechanisms with which an OAIS-type archive meets its primary mission of preserving information over the long term and making it available to the Designated Community. These mechanisms are summarized by the OAIS functional model, which defines six high-level services, or functional entities, that collectively define the OAIS's preservation and access operations: Ingest, Archival Storage, Data Management, Preservation Planning, Access, and Administration. Operating alongside these six functional entities are Common Services, which consist of basic computing and networking resources. An OAIS-type archive will implement each of the six functional entities, along with Common Services, in the course of building a complete archival system.

The reference model provides a high-level description of the information objects managed by an OAIS-type archive. The OAIS information model is built around the concept of an information package, which consists of the object that is the focus of preservation, along with metadata necessary to support its long-term preservation, access, and understandability, bound into a single logical package. There are three important variants of the information package concept: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP).

The AIP is the information package variant which the OAIS is committed to perpetuate over the long term. Construction of the AIP begins with the Content Data Object – the information that is the focus of preservation. The Content Data Object is accompanied by Representation Information: information necessary to render and understand the bit sequences constituting the Content Data Object. The Content Data Object and its associated Representation Information are collectively known as Content Information. Long-term retention of the Content Information requires additional metadata to support and document the OAIS's preservation processes. This metadata is called Preservation Description Information, or PDI. PDI consists of five components: 1) *Reference Information*; 2) *Context Information*; 3) *Provenance Information*; 4) *Fixity Information*; and 5) *Access Rights Information*. Packaging Information binds Content Information and Preservation Description Information into a single logical package; Descriptive Information supports the discovery and retrieval of Content Information by an OAIS's Consumers.

The OAIS reference model includes a discussion of different classes of interoperability across OAIS-type archives: *independent archives*, *cooperating archives*, and *federated archives*. The reference model also notes that archives can interoperate through *shared functional areas*.

A number of initiatives have used the OAI reference model as a conceptual foundation and starting point for more focused work in digital preservation. Areas of application include, but are not limited to, 'OAI-compliant' repository architectures and systems; repository self-assessment and certification; metadata requirements for digital preservation; methods and protocols for encoding and exchanging archived information; and other OAI-related standards.

Because the reference model is a conceptual framework rather than a blueprint for concrete implementation, the meaning of 'OAI-compliant' is necessarily vague and open to interpretation. A key element in the design of OAI is its flexibility and level of abstraction: it makes no assumptions about how the concepts and models in OAI are to be implemented, and imposes no requirements concerning the technologies used to support the implementations. Despite the attendant ambiguity, the notion of OAI conformance has been beneficial, to the extent that it helps consolidate understanding of the fundamental requirements for securing the long-term persistence of digital materials – a necessary condition for building well-understood, interoperable, and ultimately, trusted digital preservation systems.

Perhaps the most important achievement of the OAI reference model to date is that it has become almost universally accepted as the *lingua franca* of digital preservation, shaping and sustaining conversations about digital preservation across disparate domains, and supplying a general mapping of the landscape that stewards of our digital heritage must navigate in order to secure the long-term availability of digital materials. Alignment with concepts defined in OAI helps orient a technical implementation, draft standard, or other activity within the broader repository context that the OAI reference model defines, making it part of a cohesive 'big picture'. It seems reasonable to conclude that OAI has become a foundation resource for understanding digital preservation, a language for talking about digital preservation issues, and a starting point for implementing digital preservation solutions.

It is possible to identify a few limitations associated with OAI's impact. Very few of its concepts have been directly and formally operationalized as standards in their own right. A design, a protocol, even a standard can self-declare itself OAI-conformant (but without an explicit accounting of how conformance is actually manifested). Initiatives can use OAI concepts as a means of labelling or describing various components within their structure (but these concepts can be used quite superficially, more as an expositional shorthand rather than a detailed mapping); OAI can be cited as a foundation or starting point for a particular initiative, or alternatively the initiative can declare itself informed by OAI (but without necessarily any elaboration on how this was so). It is useful to remember that an OAI-type archive is still one built primarily on OAI *concepts*, not an OAI suite of *standards*. The digital preservation community would benefit from a careful assessment of where more precise and authoritative definitions of OAI concepts and relationships would accelerate progress in achieving robust, widely applicable, and interoperable digital preservation solutions.

3. Introduction

The impact of digital information has been remarkably universal, extending to industry, government, and the academy; to business people, scientists, engineers, and scholars of the humanities; to the individual in the workplace and the individual in the home. Vast quantities of information in digital form – text, images, audio, video, web pages, computer programs, databases – are produced, exchanged, and used in a variety of settings, for myriad purposes. These diverse applications of digital technology rest on a common foundation of shared benefits, including powerful search and retrieval capabilities, network delivery, perfect duplication, and interoperability.

The capacity both to create and consume digital information has progressed steadily, and with these advances comes the need to develop sufficient capacity to assure appropriate stewardship of this information over the long term. Technical challenges, like the relative fragility of some digital storage media, or the eventual obsolescence of storage and rendering environments, are obvious obstacles to achieving long-term preservation and access goals for digital materials, but equally important are the organizational and economic issues related to long-term digital stewardship, including allocation of and commitment to preservation responsibilities, incentives to preserve, and distribution of costs. Just as the benefits of digital information transcend people, systems, and domains, so do the challenges which accompany them.

Shared challenges invite shared solutions. The growing corpus of digital materials has confronted organizations of all descriptions – cultural heritage institutions, businesses, government agencies, etc. – with the need to take steps to secure the long-term viability of those materials in their custody. While each domain or context exhibits its own special features and circumstances, the ubiquity of the digital preservation issue establishes common ground for cross-domain dialog and cooperation in addressing the challenges of digital preservation. Moreover, efforts to develop digital preservation solutions in one community often produce a ripple effect impacting on a host of apparently unrelated communities. So it was when the space data community began to think about its own digital preservation problem: the development and subsequent release of the Open Archival Information System (OAIS) Reference Model was the progenitor of what would eventually become, and remains, one of the core standards in the broader digital preservation community.

This *Technology Watch Report* is the second edition of a report published ten years ago. While the passage of a decade has resulted in a modest addition to the first edition's story of the development of the OAIS reference model, and by extension, the contents of the model itself, the primary opportunity offered by a second edition is to reflect on the impact of OAIS by looking backward on what has been accomplished, rather than by looking forward toward what may be possible. The first edition was necessarily constrained by OAIS's newness to adopt a forward-looking focus on the reference model's potential. In this second edition, however, much more scope is available to examine how potential has been translated into impact. In addition to tracing the history of the OIAS reference model's development, and providing an overview of its key components, this report will consider the major ways OAIS has impacted on the development of digital preservation solutions. In doing so, the goal is to arrive at a general assessment of OAIS's 'legacy' at this point in its history: what were its greatest benefits? In what ways did it fall short of expectations? Of course, such an assessment is only temporary; more remains to be written as OAIS continues to develop both in form, application, and influence, and the results of a similar assessment produced for a third edition, should it be written, might be very different.

4. Development of the OAIS Reference Model

The Consultative Committee for Space Data Systems (CCSDS), established in 1982, is a forum for national space agencies interested in the cooperative development of data handling standards in support of space research.¹ In 1990, the CCSDS launched a cooperative arrangement with the International Organization for Standardization (ISO; specifically, ISO's Subcommittee 13 (space data and information transfer systems) under Technical Committee 20 (aircraft and space vehicles)) whereby CCSDS Recommendations – i.e., recommended solutions to the data-handling problems shared by its membership – would undergo normal ISO review and voting procedures in the course of becoming formal ISO Standards.

At the request of ISO, CCSDS initiated work aimed at developing formal Standards for the long-term storage of digital data generated from space missions. In preparing for this effort, CCSDS found no single widely accepted framework that could serve as a foundation for standards-building activities: nothing, for example, that established shared concepts and terminology associated with digital preservation; characterized the basic functions constituting a digital archiving system; or defined the important attributes of the digital information objects towards which preservation efforts would be directed. In short, there was no perceived consensus on the needs and requirements for maintaining digital information over the long term. A unifying framework that could fill this gap would be invaluable in terms of encouraging dialogue and collaboration among participants in standards-building activities, as well as identifying areas most likely to benefit from standards development.

In the absence of a single framework, CCSDS determined that its first step should be to create one. An international workshop convened by CCSDS in 1995² validated this strategy, and a proposal was advanced to develop a reference model for an 'open archival information system'. According to Wikipedia, a reference model 'is an abstract framework or domain-specific ontology consisting of an interlinked set of clearly defined concepts produced by an expert or body of experts in order to encourage clear communication.'³ In line with this definition, CCSDS's reference model would define the basic functional components of a system dedicated to the long-term preservation of digital information, detail the key internal and external system interfaces, and characterize the information objects managed by the system. These descriptions would be expressed in terms of a well-defined set of concepts and terminology transcending, yet mappable to, domain-specific vocabularies. The reference model would also enumerate a set of minimum requirements an archival system is expected to meet. When complete, the reference model would represent a comprehensive and consistent framework for describing and analysing digital preservation issues, provide a sound footing for future standards-building activity, and serve as a point of reference for vendors interested in building digital preservation products and services.

It should be noted that, strictly speaking, the reference model could be applied to the long-term preservation of items in any form, including physical artifacts. The reference model makes no assumptions about the nature of the information being preserved; consequently, the information could be (as the reference model itself points out) a moon rock. However, it is in the digital realm that the OAIS reference model has gained its widest visibility and acceptance, and it is within this context that it is discussed in this report.

From the earliest stages of the reference model's development, CCSDS recognized that its relevance extended well beyond the space data community. The reference model would address fundamental questions regarding the long-term preservation of digital materials that cut across domain-specific implementations. Consequently, the decision was made to make the process of crafting the model open to any interested individual or organization. In adopting this ecumenical approach, CCSDS reached beyond the space data community to engage a diverse collection of organizations in government, private industry, and academia. Developing the reference model was an opportunity to consolidate understanding of the needs

¹ For more information on the CCSDS, see <http://www.ccsds.org/>.

² Proceedings from this workshop are available at: <http://nssdc.gsfc.nasa.gov/nost/isoas/int01/ws.html>.

³ http://en.wikipedia.org/wiki/Reference_model

and requirements of digital preservation, by gathering the strands of isolated digital preservation activities into a shared characterization of the problem's boundaries.

The reference model was developed through an open, iterative process of drafting, review, and revision; community feedback was provided through face-to-face workshop discussions, and as written responses to formal requests for comment. Draft versions of the reference model were released for review in May 1997 and May 1999; it was approved and published as a draft ISO Standard in June 2000. After a final period of review and revision, the reference model was approved in January 2002 as ISO International Standard 14721, and was officially published as such in 2003.

In accordance with both ISO and CCSDS policies, the OAIS reference model underwent a review process beginning in 2006 which included a solicitation for comments from the community on perceived weaknesses of the model, needed clarifications, or portions of the model that may have passed into obsolescence.⁴ This process resulted in the circulation of a revised draft of the reference model in 2009, followed by the publication of a finalized version in 2012. This version now stands (at the time of writing) as the current edition of the OAIS reference model, approved as a CCSDS Recommended Practice, and as ISO Standard 14721:2012.⁵ With a few exceptions, the revisions are relatively modest in scale. Barbara Siermans has produced a useful summary of the revisions included in the 2012 version of the reference model;⁶ these can be briefly summarized as follows:

- Access Rights information added to Preservation Description Information;
- More emphasis on emulation as a viable preservation strategy;
- More interaction between Administration Functional Entity and Preservation Planning Functional Entity;
- 'Authenticity' defined, and linked to evidence-based assessment;
- Definition of 'Information Package' updated;
- Concept of 'Other Representation Information' introduced;
- Chapter on Preservation Perspectives updated, incorporating above revisions and updating definitions of AIP version and AIP edition.

⁴ The details of the review process can be found at <http://nssdc.gsfc.nasa.gov/nost/isoas/oais-rm-review.html>.

⁵ Early draft versions of the OAIS reference model are available at http://nssdc.gsfc.nasa.gov/nost/isoas/ref_model.htm. The first approved version of OAIS (CCSDS 650.0-B-1 [Blue Book], equivalent to ISO 14721:2003) is also available at this link. The current version of OAIS is available at <http://public.ccsds.org/publications/archive/650x0m2.pdf> (CCSDS 650.0-M-2 [Magenta Book]) or at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57284 (ISO 14721:2012). Note that the 2003 version of the reference model was designated a CCSDS Blue Book, or Recommended Standard. Subsequently, CCSDS determined that reference models would be designated as CCSDS Magenta Books, or Recommended Practice. Hence, the 2012 version of OAIS is a Recommended Practice, or Magenta Book.

⁶ See Barbara Sierman's blog post at <http://digitalpreservation.nl/seeds/oais-2012-update/>.

5. Overview of the OAIS Reference Model⁷

5.1. Open Archival Information System

The central concept in the reference model is that of an open archival information system (OAIS). The term *open* refers to the fact that the reference model was developed and released in open public forums, in which any interested party was encouraged to participate. It does not refer to, or make any suppositions about, the level of accessibility associated with an archive. An *archival information system* is ‘an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community’ (OAIS, 2012, 1-1). This definition emphasizes two primary functions for an OAIS-type archival repository: first, to preserve information – i.e., to secure its long-term persistence – and second, to provide access to the archived information, in a manner consistent with the needs of the archive’s primary users, or Designated Community. The concept of a Designated Community will be discussed in the next section.

On the surface, this definition does little to distinguish the reference model’s use of the term ‘archive’ from its usage in other contexts. However, the definition is supplemented with a list of mandatory responsibilities that an OAIS-type archive is expected to meet. In particular, an OAIS must:

- Negotiate for and accept appropriate information from information producers;
- Obtain sufficient control of the information in order to meet long-term preservation objectives;
- Determine the scope of the archive’s user community;
- Ensure that the preserved information is independently understandable to the user community, in the sense that the information can be understood by users without the assistance of the information producer;
- Follow documented policies and procedures to ensure the information is preserved against all reasonable contingencies, and that there are no *ad hoc* deletions.
- Make the preserved information available to the user community, and enable dissemination of authenticated copies of the preserved information in its original form, or in a form traceable to the original.

The responsibilities listed above are paraphrased for brevity, and edited to avoid the use of terminology that has not yet been introduced in the present discussion. The responsibilities in their original form can be found in the reference model documentation (OAIS, 2012, 3-1).

The first responsibility of an OAIS-type archive is to establish explicit selection criteria for determining which materials are appropriate for inclusion in the archival store. These criteria might be based on such factors as subject, origin, or format. Once the scope of the archival collection is defined, appropriate steps must be taken to motivate the producers/owners of the targeted items to transfer them, along with accompanying metadata, into the custody of the OAIS for preservation.

But it is not enough simply to acquire custody of the items. The second responsibility emphasizes the need for the OAIS to obtain sufficient intellectual property rights, along with custody of the items, to authorize the procedures necessary to meet preservation objectives. For example, if the OAIS must create a new version of the archived item so that it can be rendered by current technologies, it must have the explicit right to do so. The reference model notes three areas in particular where impediments may exist to obtaining needed levels of control over archived materials: 1) copyright and other legal restrictions; 2) authority to modify archived information; and 3) agreements with other organizations to share or leverage their preservation activities.

⁷ The following overview of the OAIS reference model is based on CCSDS 650.0-M-2 (Magenta Book), published in June 2012.

Another responsibility of an OAIS-type archive is to determine the scope of its primary user community. This will be discussed in greater detail in the next section, but the key point is that an accurate characterization of the primary users of the archived information is a pre-condition for meeting another of the OAIS's responsibilities: ensuring that the information is preserved in a form that is independently understandable to these users. The production of information always occurs in some context, and it is often the case that understanding this context is necessary to fully understand the information itself. Given this, the OAIS must not only preserve information, but also a sufficient portion of its associated context to ensure that the information is understandable, and ultimately, useable by future generations. 'Contextual information' that might be preserved includes, but is not limited to, a description of the structure or format in which the information is stored, explanations of how and why the information was created, and even its appropriate interpretation. Delineating the scope of the primary user community is essential for determining how much of this context should be preserved along with the information itself. This in turn has important implications for the metadata requirements needed to support the archived information.

The reference model notes that the scope of the user community can expand over time, often moving beyond narrow groups of specialists with deep, expert knowledge of the archived materials to include more general user communities. In this case, the context needed to support the understandability of archived materials with which users may have only superficial familiarity also expands. Consequently, an OAIS-type archive's view of its user communities and the context necessary to support them must be revisited – and possibly adjusted – over time.

The final two OAIS responsibilities concern the preservation process, and the mechanisms for making the archived information available to the user community. In regard to the former, an OAIS should establish and document clear policies and procedures for carrying out the preservation of the information in its custody; these should be accessible to and understandable by stakeholders in the OAIS, as well as in conformance with a set of clearly defined preservation objectives. In addition to this framework of operating procedures, the reference model also requires an OAIS-type archive to develop a clear plan that covers the disposition of the materials in the archive's custody in the event that the archive can no longer meet its preservation objectives.

Finally, an OAIS should be committed to making the contents of its archival store available to its intended user community, through the implementation of access mechanisms and services which support, to the fullest extent possible, users' needs and requirements. For example, these requirements may include preferred medium (e.g., print-on-demand, file formats) and access channels (e.g., Web access, transfer of physical media). Access restrictions attached to some or all of the archive's contents should be clearly documented.

In summary, use of the term *OAIS*, or equivalently, the term *archive* in the OAIS context, implies an archival system dedicated to preserving digital information and making it available over the long term, as well as meeting, in some form, the six mandatory responsibilities described above.

The OAIS reference model consists of three separate but related parts, each centered around the concept of an OAIS-type archive:

- The external environment within which the OAIS operates;
- The functional components, or internal mechanisms, which collectively fulfill the OAIS's preservation responsibilities;
- The information objects which are ingested, managed, and disseminated by the OAIS.

The next three sections discuss each of these parts in turn.

5.2. OAIS Environment

An OAIS-type archive does not operate in a vacuum; it carries out its preservation and access responsibilities in an environment populated by key external stakeholders with which it must cooperate. The reference model identifies and describes the external entities constituting an OAIS's environment, and characterizes the interfaces between these entities and the OAIS.

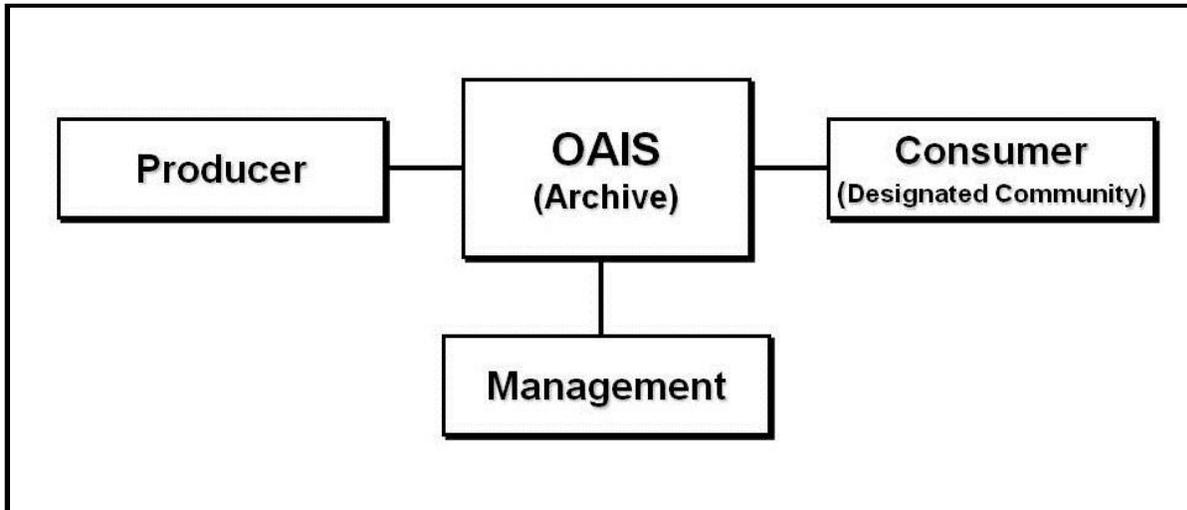


Figure 1: OAIS Environment

Figure 1 illustrates the OAIS environment. It is made up of four distinct components, three of which are explicitly external to the OAIS archive: Management, Producer, and Consumer.

5.2.1. Management

Management's responsibilities include formulating, revising, and in some circumstances, enforcing, the high-level policy framework governing the OAIS's activities. Examples of functions carried out by Management include strategic planning, defining the scope of the OAIS's archived collection, and articulating the preservation guarantee associated with items entrusted to the archive. Management may also represent the funding source for the OAIS, and often serves in an oversight capacity, periodically reviewing the OAIS's policies, performance, and risks.

It should be noted that Management is not responsible for managing the day-to-day operations of the OAIS. This responsibility is handled by a functional component within the archive itself (see section 5.3).

5.2.2. Producers

Producers are the individuals, organizations, or systems that transfer information to the OAIS for long-term preservation. Negotiation with the OAIS specifies the content and associated metadata the Producer will supply, which is then submitted to the OAIS via an ingest process that accepts the submitted data and prepares it for inclusion in the archival store. Interaction between the OAIS and Producers is often formalized and guided by a Submission Agreement, which establishes specific details of the interaction such as the type of information submitted, the metadata the Producer is expected to provide, the logistics of the actual transfer of custody from the Producer to the archive, and any access restrictions attached to the submitted material. The PAIMAS and PAIS standards, which are discussed in section 6.5, are an attempt to formally describe the interaction between Producers and OAIS archives.

5.2.3. Consumers and the Designated Community

As the name suggests, Consumers are the individuals, organizations, or systems that consume, or use, the information preserved by the OAIS. Consumers interact with an OAIS-type archive in a variety of ways, including queries for assistance, searches, and requests for access to archived information objects.

The reference model defines a special class of Consumers known as the Designated Community: the subset of Consumers expected to independently understand the archived information in the form in which it is preserved and made available by the OAIS. This point was touched on briefly in the previous section: one of the mandatory responsibilities of an OAIS is to preserve information in such a way that it is independently understandable to its primary users. These primary users are the OAIS's Designated Community.

If the OAIS contains scholarly papers and data sets specific to a particular discipline, then the Designated Community might consist of all individuals possessing a certain level of expertise in that discipline, who would use the archived information to inform and motivate basic or applied research. Similarly, if the OAIS's archived content consists of balance sheets, tax returns, and other financial records pertaining to commercial enterprises, the Designated Community might be government regulatory bodies, accountants, and other financial professionals skilled at synthesizing and interpreting this information. In both of these examples, the contents of the OAIS may be freely available for use by anyone; in this case, the OAIS's Consumers would be the general public. But it is only those individuals possessing sufficient specialized knowledge to use the archived information without expert assistance who comprise the OAIS's Designated Community.

It should not be inferred that the scope of the Designated Community is determined *ex post* by the nature of the archive's contents; rather, it is the scope of the Designated Community that determines both the contents of the OAIS and the forms in which the contents are preserved, such that they remain available to, and independently understandable by, the Designated Community.⁸

Determining the scope of the Designated Community is a critical aspect of the preservation process for an OAIS-type archive. As the discussion in section 5.4 will make clear, the broader the scope of the Designated Community, the greater the metadata requirements necessary to maintain digital materials over the long term. The Designated Community could extend as far as the public at large, which is tantamount to assuming no particular expertise or specialized knowledge on the part of users of the archived information. But in this case, the task of preserving the information in an 'independently understandable' form becomes commensurately more difficult. One additional point to note is that the scope of the Designated Community is not necessarily static: there is nothing to preclude the Designated Community from changing over time. Dynamic features of the Designated Community include its extent, as well as the expectations of its members in regard to access and use of the OAIS's contents.

5.2.4. More on the OAIS Environment Model

Although not explicitly depicted in the figure, it should be noted that an OAIS-type archive's external environment could also include interaction with other OAIS archives. For example, an OAIS could serve as a Producer in another OAIS's environment, through an agreement that transfers custody of certain information objects from the first OAIS to the second. Similarly, an OAIS may take on a Consumer role in relation to another OAIS, if it relies on the other OAIS to archive certain materials of interest to its Designated Community, and serves as an intermediary for requests for these materials.

The concepts of Management, Producers, Consumers, and Designated Community, as well as that of an OAIS, represent functional rather than organizational roles. Consequently, all of these roles can be subsumed within a single organizational structure, or distributed across multiple organizations. The key point is not the

⁸ The author thanks an anonymous reviewer for this point.

physical separation of one role from another, but rather, the logical separation of the decision-making roles and stakeholder interests attached to most digital preservation activities.

A useful example for understanding the application of the OAIS environment in practice is the Inter-university Consortium for Political and Social Research (ICPSR).⁹ This is an international organization of more than 700 institutions that ‘provides leadership and training in data access, curation, and methods of analysis for the social science research community.’¹⁰ In this context, the OAIS is ICPSR’s data archive service, which collects, curates, preserves, and makes available research data sets in the social sciences. The Management role, however, resides with the ICPSR consortium itself, which is governed by an elected council that defines organization objectives and policies. The Producers are the individuals and organizations that deposit data sets into the ICPSR data archive service, along with accompanying metadata and documentation. ICPSR’s Designated Community, or primary users, appears to consist of ‘faculty, staff and students at ICPSR member institutions who require the use of data related to the social sciences for research and instructional activities.’¹¹ However, ICPSR goes on to define its Consumer cohort more broadly to include (1) faculty, students and staff at universities and colleges that are not members of ICPSR and (2) other researchers, policymakers, service providers, and journalists at non-member institutions. Data deposit requirements call for submitted data to ‘include all data and documentation necessary to independently read and interpret the data collection’¹² – in other words, to include all supplementary information needed to make the data ‘independently understandable’ by the Designated Community. Indeed, the ICPSR web site takes the issue of ‘independent understandability’ one step further: it offers an embedded translator that will present the site’s contents in the visitor’s choice of more than 80 languages!

5.3. OAIS Functional Model

The reference model identifies and describes the core set of mechanisms with which an OAIS-type archive meets its primary mission of preserving information over the long-term and making it available to the Designated Community. These mechanisms are summarized by the OAIS functional model: a collection of six high-level services, or functional entities, that, taken together, fulfill the OAIS’s dual role of preserving and providing access to the information in its custody. The OAIS functional entities can be implemented and configured in any way appropriate to an archive’s particular circumstances and technology, but the general processes they represent should be ‘abstractable’, in one form or another, from the archive’s systems. Figure 2 illustrates the six components of the OAIS functional model.

⁹ <http://www.icpsr.umich.edu/>

¹⁰ <http://www.icpsr.umich.edu/icpsrweb/content/membership/about.html>

¹¹ <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/policies/colldev.html>

¹² <http://www.icpsr.umich.edu/icpsrweb/deposit/index.jsp>

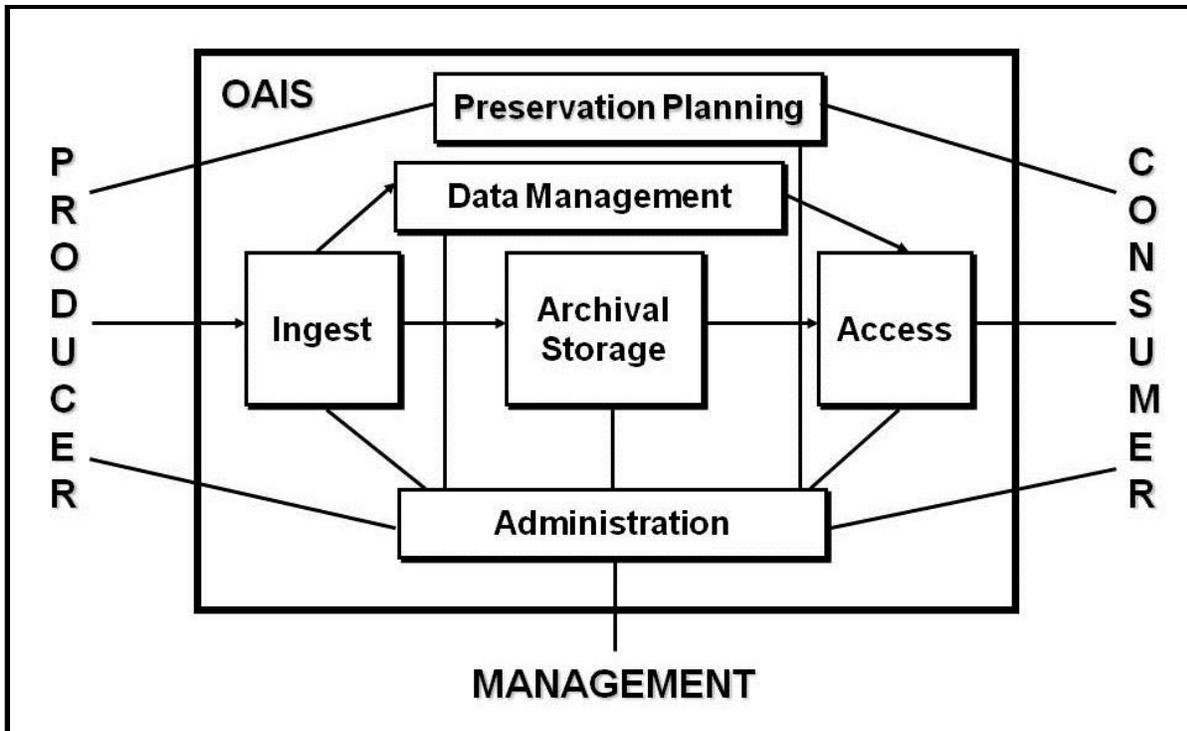


Figure 2: OAIS Functional Model

5.3.1. Ingest

Ingest is the set of processes responsible for accepting information submitted by Producers and preparing it for inclusion in the archival store. Specific functions performed by Ingest includes receipt of information transferred to the OAIS by a Producer; validation that the information received is uncorrupted and complete; transformation of the submitted information into a form suitable for storage and management within the archival system; extraction and/or creation of descriptive metadata to support the OAIS's search and retrieval tools and finding aids; and transfer of the submitted information and its associated metadata to the archival store. In short, the Ingest function serves as the OAIS's external interface with Producers, managing the entire process of accepting custody of submitted information and preparing it for archival retention.

5.3.2. Archival Storage

Archival Storage is the portion of the archival system that manages the long-term storage and maintenance of digital materials entrusted to the OAIS. More specifically, the Archival Storage function is responsible for ensuring that archived content resides in appropriate forms of storage – e.g., online, near-line, offline – and that the bit streams comprising the preserved information remain complete and renderable over the long term. To meet this responsibility, Archival Storage periodically undertakes procedures such as media refreshment or format migration. The Archival Storage function also implements various safeguard mechanisms, such as error-checking procedures, to evaluate the outcome of preservation processes, as well as disaster recovery policies to mitigate the effects of catastrophic events. Finally, Archival Storage retrieves items from the OAIS's storage systems in support of access requests by Consumers. Note that the Archival Storage function has no direct external interface; interaction with Archival Storage is confined to the OAIS's internal high-level services.

5.3.3. Data Management

The Data Management function maintains databases of descriptive metadata identifying and describing the archived information in support of the OAI's finding aids; it also manages the administrative data supporting the OAI's internal system operations, such as system performance data or access statistics. The primary functions of Data Management include maintaining the databases for which it is responsible; performing queries on these databases and generating reports in response to requests from other functional entities within the OAI; and conducting updates to the databases as new information arrives, or existing information is modified or deleted. In managing these databases, the Data Management function supports search and retrieval of the OAI's archived content, and administration of the OAI's internal operations.

5.3.4. Preservation Planning

Preservation Planning is responsible for mapping out the OAI's preservation strategy, as well as recommending appropriate revisions to this strategy in response to evolving conditions in the OAI environment. The Preservation Planning service monitors the external environment for changes and risks that could impact the OAI's ability to preserve and maintain access to the information in its custody, such as innovations in storage and access technologies, or shifts in the scope or expectations of the Designated Community. Preservation Planning then develops recommendations for updating the OAI's policies and procedures to accommodate these changes. The Preservation Planning function represents the OAI's safeguard against a constantly evolving user and technology environment. It detects changes or risks impacting the OAI's ability to meet its responsibilities, designs strategies for addressing them, and assists in the implementation of these strategies within the archival system.

5.3.5. Access

As its name suggests, the Access function manages the processes and services by which Consumers – and especially the Designated Community – locate, request, and receive delivery of items residing in the OAI's archival store. Typical services provided by Access in support of the Consumer include processing queries of the OAI's holdings – specifically, forwarding the request to Data Management and presenting the response (e.g., a result set) to the Consumer; and coordinating the retrieval and delivery of requested content – by forwarding the request to Archival Storage, receiving the requested items, and performing any necessary transformations (such as altering the archived item's format to one more suitable for dissemination, or stripping away unneeded metadata) that must occur prior to delivery to the Consumer. Access is also responsible for implementing any security or access control mechanisms associated with the archived content. The Access function represents the OAI's interface with its Consumers (and Designated Community): as such, it is the primary mechanism by which the OAI meets its responsibility to make its archived information available to the user community.

5.3.6. Administration

The Administration function is responsible for managing the day-to-day operations of the OAI, as well as coordinating the activities of the other five high-level OAI functional entities. Other responsibilities include interacting with Producers (e.g., negotiating Submission Agreements), Consumers (e.g., providing customer service support), and Management (e.g., implementing and maintaining archive policies and standards). The Administration function is also responsible for overseeing the operation of the archiving and access systems, monitoring system performance, and coordinating updates to the system as appropriate. Administration serves as the central hub for the OAI's internal and external interactions: it communicates directly with the five other OAI high-level services – Ingest, Archival Storage, Data Management, Preservation Planning, and Access, as well as the OAI's external stakeholders – Producers, Consumers, and Management.

5.3.7. More on the OAIS Functional Model

Although not shown in the figure, the OAIS reference model defines a seventh functional entity: Common Services. These services are pervasive throughout the archive, and include (among others), operating system services (e.g., basic computing resources, file management utilities); network services (e.g., data communication mechanisms); and security services (e.g., authentication/authorization). Common Services are the computing and networking backbone of any OAIS-type archive.

In summary, an OAIS is expected to embody six high-level functional entities that, together with Common Services, constitute the mechanisms by which the OAIS preserves information over the long-term and makes it available to its Designated Community. An OAIS-type archive will implement each of these functional entities, in one form or another, in the course of building a complete archival system.

The OCLC Digital Archive¹³ service provides an illustration of how the six OAIS high-level services might be implemented in practice. The architecture of this service is based on the OAIS reference model: therefore, each component of the OAIS functional model is recoverable from the wide array of OCLC organizational units supporting the Digital Archive. Specifically, the OCLC Digital Archive draws on OCLC Digital Collection Services (Ingest); Global Systems and Information Technology (Archival Storage); Database Support Systems (Data Management); OCLC Research, Systems Planning (Preservation Planning); End-user Services, Digital Collection Services (Access); and Corporate Security, Legal, Network Support, Systems Support (Administration). The OCLC Digital Archive exemplifies the diverse collection of processes, services, and expertise that must be integrated to produce the six OAIS high-level functional components. However, the OCLC approach is but one of many possible strategies for implementing the OAIS functional model.

5.4. OAIS Information Model

In addition to describing the functional components of an OAIS-type archive, the reference model also provides a high-level description of the information objects managed by the archive. The OAIS information model is built around the concept of an information package: a conceptualization of the structure of information as it moves into, through, and out of the archival system. An information package consists of the object that is the focus of preservation, along with metadata necessary to support its long-term preservation, access, and understandability, bound into a single logical package. There are three important variants of the information package concept: the Submission Information Package, the Archival Information Package, and the Dissemination Information Package.

5.4.1. Information Packages

The Submission Information Package, or SIP, is the version of the information package that is transferred from the Producer to the OAIS when information is ingested into the archive. The exact form of the SIP may be the result of a negotiated agreement between the Producer and the OAIS, or it may be constructed on an *ad hoc* basis: e.g., the digital object and as much metadata as the Producer is willing or able to supply. The concept of the SIP emphasizes the fact that information may not be preserved in the exact form in which it is submitted by the Producer. For example, the preserved object may be the aggregation of content provided in multiple SIPs; or, the Producer may provide the information in a format not supported by the OAIS, necessitating migration to another format prior to inclusion in the archival store. It may also be the case that the metadata supplied by the Producer is incomplete or inadequate, and must be augmented during the ingest process.

The Archival Information Package, or AIP, is the version of the information package that is stored and preserved by the OAIS. The AIP consists of the information that is the focus of preservation, accompanied by a complete set of metadata sufficient to support the OAIS's preservation and access services. The archived

¹³ <http://oclc.org/en-US/digital-archive.html>

information and its associated metadata represent a single logical package within the archival system: there is, however, no requirement that any form of physical association be maintained, such as embedding the metadata in the information object itself and storing the combined object as a single bit stream. Arrangements for storing archived information and its metadata are left to the OAIS's implementers; possible solutions might range from complete physical integration, to storage in separate yet logically related databases.

The reference model defines two 'specializations' of an AIP: an Archival Information Unit (AIU), and an Archival Information Collection (AIC). An AIU stores the content and metadata for a single 'atomistic' object (e.g., a single digital movie or e-book), while an AIC consists of multiple AIUs that have been grouped together into a distinct collection. Note that the unitary or atomistic nature of the object associated with an AIU is conceptual: in practice, the object could exist as multiple physical or digital parts (for example, each chapter of an e-book could be stored in a separate file). In the case of an AIC, each of the component AIUs, as well as the AIC itself, will have its own associated metadata. A single AIU can be part of multiple AICs; in addition, an AIC itself can be a component of a broader AIC. AICs may group AIUs together based on shared subject, theme, origin, or any other criteria that suits the purpose of the OAIS in which they reside. The reference model notes, for example, that AICs may streamline the process of searching an archive's contents by organizing individual AIUs into useful hierarchies. (OAIS, 2012, 4-45) AICs may also be of use to the archive itself in its day-to-day operations: for example, an AIC may group together objects requiring a specific preservation technique. In short, AICs are an organizational device in the form of a conceptual layer (i.e., metadata describing the AICs) that sits on top of individual AIUs in an OAIS-type archive.

Finally, the Dissemination Information Package, or DIP, is the version of the information package delivered to the Consumer in response to an access request. The DIP concept emphasizes the fact that the information package disseminated by the OAIS to the Consumer may differ in form or content to that which resides in the archival store. Points of differentiation between the DIP and AIP may include, but are not limited to, the format of the content (e.g., an image file might be converted from TIFF to JPEG prior to dissemination); the amount of content (a DIP may correspond to one AIP, multiple AIPs, or even part of an AIP); and the amount of metadata supplied alongside the content (it is likely that the DIP will not contain the complete set of metadata associated with an archived digital object, since much of it is of little interest to the Consumer).

SIPs, AIPs, and DIPs represent the information objects deposited into, managed by, and disseminated from an OAIS-type archive. But it is the AIP – the Archival Information Package – that is the focus of preservation: it is the information package variant which the OAIS is committed to perpetuate over the long term. Given the importance of the AIP in regard to the OAIS's preservation and access responsibilities, it is useful to take a closer look at this information package and examine its key components.

5.4.2. Inside the Archival Information Package

Recall that an information package contains the content to be preserved, along with its associated metadata. The AIP embodies a stricter interpretation of this concept, in that it must include the complete set of metadata necessary to support the content's long-term preservation and availability to the Designated Community. The reference model characterizes the types of metadata that should be included with the archived information. Figure 3 illustrates the information components of an AIP.

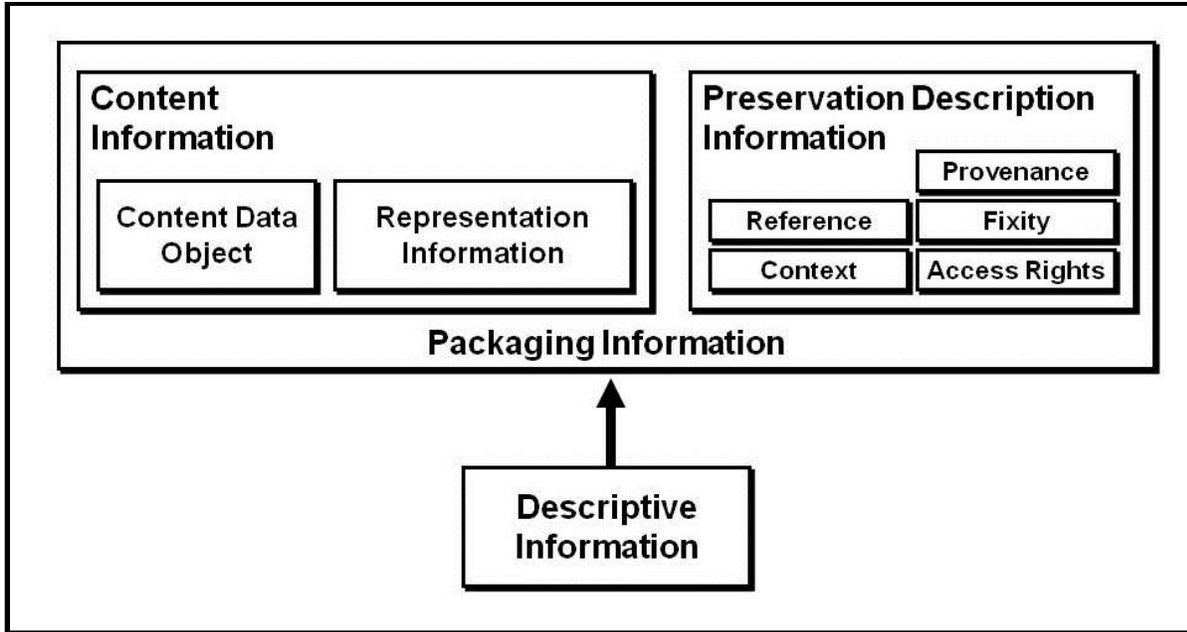


Figure 3: Archival Information Package

Content Data Object

Construction of an AIP begins with the Content Data Object – the information that is the focus of preservation. The Content Data Object can take the form of any class of material: text, images, video, databases, computer programs – even physical material such as soil samples or fossils. The Content Data Object may be comprised of a single, self-contained object – for example, a document in PDF format; it may also encompass multiple objects, such as a website consisting of text (HTML files) and static images (GIF or JPEG files). The key point is that the OAIS is responsible for preserving the Content Data Object over the long term, as well as for making it available in a form that is independently understandable by the Designated Community.

Representation Information

In order to meet the second responsibility – to make the Content Data Object available in a form that is independently understandable by the Designated Community – the Content Data Object must be accompanied by an appropriate quantity of Representation Information: information necessary to render and understand the bit sequences constituting the Content Data Object. Representation Information might include a description of the hardware and software environment needed to display the Content Data Object and/or access its contents; it might also summarize the appropriate interpretation of the Content Data Object. For example, if the Content Data Object is an ASCII file of numbers, Representation Information might indicate that the numbers correspond to average daily air temperature readings for London, measured in degrees Celsius, for the period 1972 – 2000.

Representation information can be divided into two types: Structure Information and Semantic Information. Structure Information is most easily understood in the context of digital objects, and refers to mappings between digital bits and various concepts and data structures that render the bits into intelligible information – i.e., an image, text, an interactive program. Generally speaking, Structure Information describes the *format* of the digital object. Semantic Information, on the other hand, is information that clarifies the meaning or appropriate interpretation of the Content Data Object. A glossary, a data dictionary, and a software application's user documentation are all examples of Semantic Information that may be bundled with the Content Data Object as part of its Representation Information. The reference model also defines a catch-all category called Other Representation Information, which includes any Representation Information that is not

easily defined as either Structure or Semantic. For example, the reference model notes that information that explicates how the Structure and Semantic Information relate to one another would fall into this category.

In practice, the structure of Representation Information can be extremely complex. A particular set of Representation Information may require additional Representation Information in order to be rendered, interpreted, and/or understood by the Designated Community. The second set of Representation Information may itself require yet another set of Representation information. This regressive process can continue for an arbitrary number of steps. For example, consider a digital object in the form of a METS (Metadata Encoding and Transmission Standard) document. To ensure understandability of a METS document, an OAIS-type archive might need to secure a copy of the METS schema as part of the object's Representation Information. However, the METS schema is expressed in XML (Extensible Markup Language); therefore, to understand the METS schema (and therefore, indirectly, to understand the original METS document), users may need access to the XML specification. XML is itself a profile of the SGML (Standard Generalized Markup Language) ISO Standard 8879:1986; therefore, to fully understand XML, a copy of the SGML standard might also be needed as part of the original object's Representation Information. All of these materials – the METS schema, the XML specification, the SGML standard – form a Representation Network associated with the Content Data Object (the METS document). Representation Networks are nested chains of information that form sufficient context for the Designated Community to understand a Content Data Object, as well as its associated Representation Information. In theory, Representation Networks can form an infinite regression leading to absurd results: continuing our METS example, one could say that the SGML standard is available as ASCII text, so a copy of the ASCII specification is needed to understand it; the ASCII specification is published in English, so an English language dictionary and grammar rules are needed to understand the ASCII specification, and so on. In practice, of course, the OAIS archive will truncate the Representation Network at an appropriate point based on reasonable assumptions about the prior or assumed knowledge possessed by the Designated Community – for example, an assumption that the Designated Community understands the English language. The OAIS reference model refers to this assumed knowledge as the Designated Community's Knowledge Base.

It was mentioned earlier that the scope of the Designated Community impacts the amount of metadata required to support the preservation process. It is in regard to Representation Information that this is so. In general, the broader the scope of the Designated Community, the less specialized the knowledge associated with that community – that is, the less information relevant to interpreting and understanding the archived information the OAIS can assume its Designated Community possesses. The less specialized the knowledge base, the more Representation Information is needed to ensure that the preserved information remains renderable and understandable to the Designated Community over the long term. In this sense, Representation Information is a significant source of risk for an OAIS-type archive: as the Designated Community evolves and possibly expands over time, the archive must ensure that the Representation Information it captures and maintains evolves accordingly. This may be a challenging task, because as Representation Information requirements expand over time, the archive might be called upon to retroactively supplement Representation Information for Content Data Objects that have been in archival retention for a considerable period of time. The ease with which such Representation Information can be obtained, or indeed whether it is still available at all, is an open question.

The Content Data Object and its associated Representation Information (or Network) are collectively known as Content Information. It is the Content Information – the information that is the focus of preservation, along with sufficient metadata to ensure it remains renderable and understandable to the Designated Community – that the OAIS must perpetuate over time.

Preservation Description Information

Long-term retention of the Content Information requires additional metadata to support and document the OAIS's preservation processes. This metadata is called Preservation Description Information, or PDI. According to the reference model, PDI 'is specifically focused on describing the past and present states of the

Content Information, ensuring that it is uniquely identifiable, and ensuring it has not been unknowingly altered' (OAIS, 2012, 4-29).

PDI consists of five components:¹⁴

- *Reference Information* uniquely identifies the Content Information within the OAIS's internal systems, as well as to entities and systems external to the OAIS. Examples include a system-generated internal identifier, and an ISBN.
- *Context Information* describes the Content Information's relationships to other Content Information objects: for example, those that are related to it thematically (e.g., as part of a subject-based collection), or those that represent versions of the same content in alternative formats.
- *Provenance Information* documents the history of the Content Information, including its creation, any alterations to its content or format over time, its chain of custody, any actions taken to preserve the Content Information (such as normalization or format migration), and the outcome of these actions.
- *Fixity Information* ensures that the Content Information has not been altered in an undocumented way, through authenticity or integrity validation mechanisms such as check sums, digital signatures, or digital watermarks.
- *Access Rights Information* documents any conditions or restrictions associated with the Content Information pertaining to both preservation and access. It may also include descriptions of rights enforcement mechanisms. Examples include licence terms, identification of those with authorized access permissions (e.g., a specified IP address range), and preservation terms and conditions negotiated between the OAIS archive and the Producer of the Content Information.

Taken together, the Content Information and Preservation Description Information represent the archived digital content, the metadata necessary to render and understand it, and the metadata necessary to support its preservation, authenticity, and dissemination.

Packaging Information is used to bind Content Information (Content Data Object and Representation Information) and Preservation Description Information (Reference, Context, Provenance, Fixity, and Access Rights Information) into a single logical package. More specifically, Packaging Information serves to combine (logically) all of these information components into an AIP, permitting them to be identified and located as a single logical unit within the archival system. Packaging Information might take the form of basic information such as directory paths and file names, or more detailed packaging schema such as METS.

Finally, Descriptive Information is information that supports the discovery and retrieval of Content Information by an OAIS's Consumers, via its finding aids. For example, Descriptive Information might take the form of a Dublin Core metadata record, derived from the Content Information and its associated Preservation Description Information, and maintained by the OAIS to facilitate discovery on the part of the archive's users.

Putting the Pieces Together

The information components described above – Content Information (Content Data Object and Representation Information), Preservation Description Information (Reference, Context, Provenance, Fixity, and Access Rights Information), Packaging Information, and Descriptive Information – collectively form the information model of an OAIS-type archive. More specifically, Content Information and Preservation Description Information form an Archival Information Package; Packaging Information allows the AIP to be identified and located as a single logical unit; and Descriptive Information supports discovery and dissemination of the AIP. Just as the OAIS reference model does not prescribe any particular approach to implementing the functional model described in section 5.3, in the same way it eschews any specific recommendation for implementing the various components of the information model. The goal is instead to

¹⁴ With the exception of Access Rights Information, the components of PDI are based on discussion in the landmark report by Waters and Garrett 1996; see p. 11–19.

provide a conceptual model of the information objects managed by an OAIS-type archive. Implementation of these concepts will depend on the specific architectures, systems, and schema employed in a particular archival setting.

5.5. Archive Interoperability

The OAIS reference model includes a discussion of interoperability across OAIS-type archives. As an increasing proportion of the scholarly and cultural record reside in digital repositories, it is likely that the network of relationships overlaid on top of these repositories will expand and become more complex. Distinct OAIS archives – that is, OAIS archives with separate Managements – can benefit from cooperation in a variety of ways, such as providing Consumer communities a more uniform, integrated access environment across archives; streamlining the submission process for Producers who deposit materials in multiple archives; and leveraging opportunities to reduce operational costs through shared infrastructure and services. The reference model supplies a taxonomy describing four possible inter-archive associations.

Independent archives are OAIS-type archives that operate autonomously, with no interactions with other archives. The activities of independent archives are focused on a single Designated Community, and decision-making within the archive is based exclusively on the needs of that Community, or the archive’s internal requirements.

Cooperating archives are two or more archives that maintain some form of submission or dissemination compatibility between them. More specifically, the archives support at least one SIP or DIP format that is used to fulfill requests made from one archive to another. For example, cooperating archives may agree on SIP and DIP formats such that a DIP from one archive can be easily ingested as a SIP by another archive. Motivations for this level of cooperative compatibility may include shared Consumer communities (who would be interested in materials held in all of the cooperating archives) and/or shared Producers (who submit materials to all of the cooperating archives). Cross-repository overlap in Consumers and Producers strengthens the incentive to reduce frictions in transferring materials from one archive to another.

Federated archives are OAIS-type archives that serve not only a ‘local’ Designated Community (defined as the original community that an archive was set up to serve), but also a ‘global community’ that is served by the materials in multiple archives. The global community is able to access the collective holdings of the federated archives through one or more shared discovery/fulfillment mechanisms. These mechanisms might take the form of a central catalogue of the collective contents of the federation, which directs Consumers to the appropriate archive for access to specific materials; a global search function that distributes Consumer search queries to all of the archives in the federation, with each archive processing the query against its local catalogue; or a fully integrated global search and access service that locates and retrieves requested material from the appropriate archive through a process transparent to the Consumer.

Finally, archives can interoperate through *shared functional areas*. In this case, the Management entity of each archive agrees to share resources (infrastructure, systems, services, etc.) used to support one or more of the six functional areas defined in the OAIS reference model (Ingest, Archival Storage, Administration, Data Management, Preservation Planning, and Access). For example, a group of archives might share a common registry of Representation Information (such as the PRONOM registry of technical metadata for software and digital formats).¹⁵ Sharing resources of this kind can reduce the cost of long-term preservation by leveraging economies of scale and minimizing unnecessary duplicative capacity.

The OAIS reference model’s discussion of various forms of cross-archive association or integration underscores the point that an important aspect of the external environment of most OAIS-type archives is the presence of other OAIS-type archives. Consequently, few OAIS-type archives will operate with complete autonomy (i.e., as independent archives), but instead will form associations of varying degrees of integration

¹⁵ <http://www.nationalarchives.gov.uk/PRONOM/>

with other archives. In these circumstances, the concepts expressed in the reference model – the OAIS environment, functional, and information models – should be considered in the context of a multi-archive setting, rather than that of a single archive operating in isolation. Careful thought should be given in the design of an OAIS-type archive to possible sources of cross-archive standardization or integration across the elements of the reference model: for example, overlap across archives’ Producer or Consumer communities; opportunities to share infrastructure and services within OAIS internal functional entities; and adoption of standards and common formats to facilitate the transfer of information packages between archives.

6. Impact of the OAIS Reference Model

OAIS is a model and not an implementation. The standard says nothing about system architectures, storage or processing technologies, database design, computing platforms, or any of the myriad technical details involved in setting up a functioning archival system. However, since OAIS's publication, a number of initiatives have used it as a building block – a conceptual foundation and starting point – in support of the construction of functioning archiving systems. This section briefly describes a number of OAIS-related activities that built on the reference model and its concepts in very practical ways. The list of activities discussed in this section is intended to be illustrative rather than exhaustive.

6.1. OAIS-compliant repository architectures

The reference model states that an OAIS-compliant archive supports the OAIS information model. It also is committed to meet the mandatory responsibilities enumerated by the reference model (OAIS, 2012, 1-3–1.4). Finally, the reference model notes that standards and other documentation that purport to conform to the OAIS reference model must incorporate relevant OAIS terminology and concepts, applied according to the interpretation and context defined in the reference model. The reference model takes pains to emphasize that OAIS conformance does not entail specific technology choices, or other constraints on implementation decisions.

In surveying the digital preservation landscape today, it is not uncommon to find references to the OAIS reference model in descriptions of digital archiving solutions. For example, the Rosetta digital preservation system offered by Ex Libris describes itself as based on the OAIS reference model.¹⁶ Tessella's Preservica digital preservation solution highlights its suite of OAIS compliant workflows.¹⁷ Similarly, Archivematica, an open-source digital preservation system, provides 'an integrated suite of software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model.'¹⁸ The Dark Archive in the Sunshine State (DAITSS) open-source digital preservation application was 'deliberately designed to meet the requirements of an OAIS'.¹⁹ The Lots of Copies Keeps Stuff Safe (LOCKSS) system, which supports networks of libraries cooperatively preserving their digital content, produced a Formal Statement of Conformance to ISO 14721:2003²⁰ in support of their adherence to 'industry standards'.²¹ The Repository of Authentic Digital Objects (RODA) open-source repository system is 'supported by existing standards such as the OAIS' and its design is explicitly mapped to the six OAIS functional entities.²²

What validation does OAIS compliance provide to stakeholders in the long-term preservation of digital materials? Because the reference model is a conceptual framework rather than a blueprint for a concrete implementation, the meaning of OAIS-compliant is necessarily vague. Conformance to the reference model can imply an explicit application of OAIS concepts, terminology, and the functional and information models in the course of developing a digital repository's system architecture and data model; but it can also mean that the OAIS concepts and models are recoverable from the implementation – in other words, it is possible to map, at least from a high-level perspective, the various components in the archival system to the corresponding features of the reference model. Further ambiguity is introduced when institutions and organizations claim OAIS compliance without defining or clarifying what this means in regard to their particular implementation.

A key element in the design of OAIS is its flexibility and level of abstraction: it makes no assumptions about how the concepts and models in OAIS are to be implemented, and imposes no requirements concerning the

¹⁶ <http://www.exlibrisgroup.com/category/RosettaOverview>

¹⁷ <http://preservica.com/preservica/>

¹⁸ https://www.archivematica.org/wiki/Main_Page

¹⁹ http://daitss.fcla.edu/sites/daitss.fcla.edu/files/DAITSS%20in%20ACM%20rev_0.pdf

²⁰ <http://www.lockss.org/locksswp/wp-content/uploads/2011/11/OAIS-LOCKSS-Conformance.pdf>

²¹ <http://www.lockss.org/about/how-it-works/#industry-standards>

²² <http://www.roda-community.org/what-is-roda/>

technologies used to support the implementations. While these features serve to extend the reference model's applicability to almost any digital preservation scenario, they come at the price of consistency in interpretation and application of the notion of OAIS conformance. Jerome McDonough summarizes this point nicely (albeit in a different context) in noting 'the tremendous degree of play in encoding practice' for XML-based metadata schemas. McDonough concludes that 'The digital library community seems to face a dilemma at this point. Through its pursuit of design goals of flexibility, extensibility, modularity and abstraction, and its promulgation of those goals as common practice through its implementation of XML metadata standards, it has managed to substantially impede progress towards another commonly held goal, interoperability of digital library content across a range of systems' (McDonough, 2008).

In a similar way, the high-level, conceptual make-up of the OAIS reference model poses a tradeoff between flexibility on the one hand, and consistency in implementation on the other. Two examples illustrate this point. In order to be OAIS conformant, all OAIS-type archives must support the OAIS information model, but the information model can be 'supported' in a variety of ways. One repository may meet the metadata requirements of an archival information package via the semantic units defined in the PREMIS Data Dictionary; another may do so through an internally designed schema. Both archives are OAIS conformant, but in the absence of a detailed mapping between PREMIS and the internally designed schema, their respective implementations of an archival information package will not be interoperable. Similarly, all OAIS-type archives must meet the mandatory responsibilities enumerated by the reference model, but the mechanisms by which they do so are not prescribed. One repository that archives research data sets may define its Designated Community very broadly, perhaps extending it to include the general public; in these circumstances, the representation information needed to make the archived data sets understandable to the Community will likely be extensive. Another repository that archives similar kinds of data sets may view its Designated Community much more narrowly, perhaps limiting it to experts; in these circumstances, the representation information requirements will likely be comparatively less. Again, both repositories are OAIS conformant, but archival information packages from the second repository would not be equipped with sufficient information to make them understandable to the first repository's Designated Community.

This assessment of the practical implications of OAIS conformance may be disappointing to those who anticipated a much more precise meaning: e.g., a rigorous application of a well-defined suite of standards, protocols, and best practices. Yet the importance of OAIS-conformant archives should not be discounted. A shared view of the core functional and information requirements of digital archiving systems is essential for creating long-term preservation solutions that are well-understood and accepted by a potentially extended stakeholder community. For example, the mutual understanding promoted by the OAIS reference model would benefit the process of negotiating service level agreements in situations where digital preservation activities are outsourced to external providers.

Moreover, a high-level, shared view of the salient contours of a digital preservation repository serves as a starting point for, and facilitates the development of, the standards, protocols, and best practices that underpin an interoperable network of digital archives. Indeed, one of the original motivations for producing the reference model was to put forward a widely-applicable framework that would serve as a starting point for more focused standards- or consensus-building activities. As we will see below, a number of these activities have been undertaken since the release of the reference model, yielding a significant body of work that, taken together, offers a more distinct picture of a 'real-world' OAIS-type archive. As this work continues to progress, digital repositories currently incorporating nothing more than OAIS terminology and concepts will find that even this loose conformance to the reference model facilitates the process of adopting new OAIS-related standards and best practices as they become available.

The significance attached to OAIS conformance ultimately rests on whether it produces a tangible impact on stakeholder confidence in a digital repository's ability to meet its preservation objectives. Libraries, museums, and other collecting institutions, for example, are faced with the prospect of entrusting irreplaceable portions of the scholarly and cultural record to digital archiving systems whose capacity to provide effective long-term stewardship is as yet unproven. Does OAIS conformance make this decision

easier? The answer, it would seem, is ‘yes’. As we have seen, commercial and open-source digital preservation solutions see utility in claiming OAIS conformance as a selling point or feature of their systems. As we will discuss in the next section, a prominent audit and certification program for digital repositories is based on metrics which are themselves based on the OAIS reference model. And as we will discuss in subsequent sections, a number of OAIS-related standards have emerged that have become key resources within the digital preservation community. In short, the OAIS reference model has been quite successful in consolidating understanding of the fundamental requirements for securing the long-term persistence of digital materials. A shared perception of these requirements serves as a point of familiarity in what can be an uncertain landscape; it is also a necessary condition for building well-understood, interoperable, and ultimately, trusted digital preservation systems.

6.2. Repository self-assessment and certification

As repositories for digital information emerged in response to a rising need for digital preservation capacity, interest grew in establishing mechanisms for auditing, and possibly certifying, repositories’ ability to meet certain minimum requirements regarding the long-term stewardship of digital information. A natural place to start in enumerating these requirements was the OAIS reference model, given its ubiquity in the digital preservation community. In 2002, an initiative jointly sponsored by OCLC and Research Libraries Group (RLG) published *Trusted Digital Repositories: Attributes and Responsibilities*,²³ which described the attributes of a trusted digital repository. Based on the work of an international working group, the report translated OAIS concepts and models into a consensus statement on the responsibilities and characteristics of a digital repository housing a large-scale, heterogeneous collection of culturally significant materials. A key objective of this effort was to enumerate attributes of a digital repository that, taken together, would inspire trust within the Designated Community that the repository is indeed capable of preserving and making available the portion of the scholarly and cultural record in its custody.

In 2003, a joint task force sponsored by RLG and the US National Archives and Records Administration (NARA) was formed to extend the attributes of a trusted digital archive into a ‘checklist’ that could be used to support a repository certification process. The Task Force on Digital Repository Certification was charged with developing ‘criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections’ (TRAC, 2007, 2). This effort built on the concepts of the OAIS reference model, as well as the definition of a trusted digital repository described in the 2002 RLG/OCLC report. The Task Force published its final report in 2007: *Trustworthy Repositories Audit & Certification: Criteria & Checklist*, or TRAC. The TRAC checklist represents ‘best current practice and thought about the organizational and technical infrastructure required to be considered trustworthy and capable of certification’ (TRAC, 2007, 2). The scope of the checklist covers three broad areas: organization and governance; management of digital objects; and technology.

Following its initial release in 2007, TRAC was revised and published as CCSDS 652.0-M-1 (Magenta Book)²⁴ in 2011, as part of the process of becoming an ISO Standard. Subsequently, an ISO Standard (ISO 16363)²⁵ based on TRAC was approved in 2012. In parallel with TRAC’s development as an ISO Standard, a companion Standard – *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories* – was drafted, which spells out minimum requirements for the accreditation of organizations carrying out TRAC-based repository certification, as well as for the audit process itself. These requirements were published as CCSDS 652.1-M-2 (Magenta Book)²⁶ in early 2014, and at the time of writing are under review for approval as ISO 16919.

²³ <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf?urlm=161690>

²⁴ <http://public.ccsds.org/publications/archive/652x0m1.pdf>

²⁵ http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

²⁶ <http://public.ccsds.org/publications/archive/652x1m2.pdf>

The TRAC checklist addresses the difficulty noted in the previous section with the idea of OAIS conformance: as the TRAC checklist itself notes, ‘Institutions began to declare themselves ‘OAIS-compliant’ to underscore the trustworthiness of their digital repositories, but there was no established understanding of ‘OAIS-compliance’ beyond meeting the high-level responsibilities defined by the standard’ (TRAC, 2007, 1). In this sense, the TRAC checklist can be viewed as *one way* of defining an OAIS-compliant archive in concrete terms, based on well-defined and measurable criteria that can be mapped to real-world repositories organizations and systems. Put another way, TRAC unpacks and interprets in greater detail the general conformance requirements specified in the OAIS reference model (and described in the previous section). It should be noted, however, that TRAC represents only one possible interpretation of these conformance requirements.²⁷ Given the very high-level nature of the OAIS conformance requirements, it is perfectly feasible that another OAIS-based effort to standardize repository certification could arrive at a very different set of specific criteria, which could nevertheless be rolled up into the same broad conformance requirements on which TRAC is based. All such interpretations would be anchored to the same conceptual foundation (OAIS), and taken in combination, may even prove complementary. However, they would differ in the way OAIS concepts are translated into auditable characteristics of real-world archiving systems.

The Data Seal of Approval (DSA) is another OAIS-based certification available to digital repositories.²⁸ Developed by the Dutch institute DANS (Data Archiving and Networked Services) and now administered by an international governing board, the DSA is given to repositories that ‘demonstrably comply’ with 16 guidelines, many of which can be traced to elements of the OAIS reference model. For example, the DSA requires repositories to incorporate a technical infrastructure that ‘explicitly supports the tasks and functions described in internationally accepted standards like OAIS’. In addition, the DSA relates its guidelines to a stakeholder community that strongly reflects the OAIS Environment model, including such entities as Data Producers and Data Consumers.²⁹ A typical procedure for receiving DSA certification involves a self-assessment by the candidate repository, which is submitted to a peer review process for approval.

It is still an open question whether an external audit and certification process based on TRAC, or another standard, offers significant benefits over self-declared OAIS compliance. The Center for Research Libraries in the United States has performed external audits on four digital repositories using the TRAC checklist, with the results publicly available on its website.³⁰ Other repository audit tools, such as DRAMBORA³¹ and the Drupal TRAC review tool,³² are intended for internal use, or self-auditing. These too offer a range of benefits, such as signalling repository stakeholders that archiving systems are regularly evaluated according to an externally developed, standardized protocol, and cultivating workflows and other internal practices that are benchmarked for quality against well-defined assessment criteria. And as noted above, there is nothing to prevent a repository from declaring itself OAIS compliant without the aid of any formal audit or certification standard. It seems reasonable to suggest that none of these approaches are intrinsically superior to the others; ultimately, an assessment of the relative costs and benefits of each approach, combined with the specific circumstances and needs of the digital preservation setting in which they are to take place, will determine which method best suits a particular repository.

6.3. Metadata for digital preservation³³

The OAIS information model describes a broad set of metadata requirements needed to support the activities of an OAIS-type archive. These requirements have had a deep influence on the subsequent development of a number of preservation metadata schema. Preservation metadata is ‘metadata that supports the process of

²⁷ See for example, nestor’s Catalogue of Criteria for Trusted Digital Repositories (http://files.d-nb.de/nestor/materialien/nestor_mat_08-eng.pdf), which is also based on OAIS.

²⁸ <http://datasealofapproval.org/>

²⁹ <http://datasealofapproval.org/en/information/guidelines/>

³⁰ <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0>

³¹ <http://www.repositoryaudit.eu/>

³² https://www.archivematica.org/wiki/Internal_audit_tool

³³ This section is partly based on Lavoie & Gartner 2013.

long-term digital preservation’, including information about the provenance, intellectual property rights, and technical and interpretive environment of an archived digital object (Lavoie and Gartner, 2013, 4–5). In this sense, the scope of preservation metadata intersects considerably with the types of metadata defined in the OAIS information model, and therefore OAIS is a natural starting point and foundation for efforts to develop preservation metadata schema.

Although there have been a number of preservation metadata initiatives, one – PREMIS – has emerged as the *de facto* standard in this area. The origins of PREMIS can be traced to an international working group³⁴ convened by OCLC and RLG, which defined the concept of preservation metadata, and described its significance in regard to the long-term persistence of digital materials. The group also reviewed and synthesized a number of existing preservation metadata schema, with the purpose of identifying points of convergence – one of which was the use of OAIS information model concepts as a starting point for schema development. The working group published its findings in a 2001 white paper,³⁵ which became the foundation for the group’s efforts to develop a comprehensive preservation metadata framework that would identify and describe the types of information needed to support the preservation of digital materials. The framework took the form of an expanded conceptual structure for the OAIS information model, along with a set of ‘prototype’ metadata elements mapped to the conceptual structure and reflecting the information concepts and requirements set forth in the OAIS reference model. The working group published its framework in 2002.³⁶

OCLC and RLG convened a second working group in 2003 that used the framework as a foundation for developing an implementable set of core preservation metadata elements, supported by a data dictionary, and broadly applicable within the digital preservation community. Published in 2005,³⁷ the PREMIS (Preservation Metadata: Implementation Strategies) Data Dictionary is ‘a comprehensive guide to the core metadata needed to support long-term digital preservation.’ (Lavoie and Gartner, 2013, 9) In particular, it defines a collection of ‘semantic units’ – discrete pieces of information – that constitutes the metadata needed to support the preservation process in most digital repository settings. Like the OAIS reference model itself, PREMIS is implementation- and technology-neutral, although a set of XML schemas are available to support implementation.³⁸ The PREMIS Data Dictionary, schema, and other resources are hosted and maintained by the US Library of Congress.³⁹

Since its original publication in 2005, the PREMIS Data Dictionary has been revised and updated several times, most recently in 2012 with the release of PREMIS 2.2.⁴⁰ PREMIS has emerged as the *de facto* international standard for preservation metadata, and has been implemented in digital repositories worldwide. Its importance in the digital preservation community has earned it several awards, including the 2005 international Digital Preservation Award,⁴¹ and the 2006 Society of American Archivists Preservation Publication Award.⁴² In 2012, PREMIS was shortlisted for the inaugural Decennial Digital Preservation Award, which honors the most significant contribution to digital preservation in the last decade.⁴³

Although the scope of the PREMIS Data Dictionary does not overlap precisely with the metadata requirements described in the OAIS information model – for example, the OAIS concept of Descriptive Information is not covered by PREMIS – the key OAIS concepts of representation information and preservation description information are the foundation for PREMIS (as they were for the preservation

³⁴ <http://www.oclc.org/research/projects/pmwg/wg1.htm>

³⁵ http://www.oclc.org/content/dam/research/activities/pmwg/presmeta_wp.pdf

³⁶ http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf

³⁷ http://www.loc.gov/standards/premis/v1/premis-dd_1.0_2005_May.pdf

³⁸ <http://www.loc.gov/standards/premis/schemas.html>

³⁹ <http://www.loc.gov/standards/premis/>

⁴⁰ <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>

⁴¹ <http://www.dpconline.org/newsroom/not-so-new/110-awards-2005>

⁴² <http://www.oclc.org/research/news/2006/08-25b.html>

⁴³ <http://www.dpconline.org/advocacy/awards/2012-digital-preservation-awards>

metadata framework upon which PREMIS was based). Indeed, PREMIS can be viewed as an implementation of these concepts, although not necessarily one that covers the full range of information they are intended to encompass. The key point is that the OAIS information model serves as the foundation for what is now the prevailing standard for preservation metadata. In the same way that the OAIS functional model has exerted a deep influence on the development of digital preservation repository systems and services, the OAIS information model has wielded a similar influence in the design of the information objects the repositories are built to manage.

6.4. Encoding and exchanging archived information

While PREMIS addresses the metadata requirements specified in the OAIS reference model, other initiatives have focused on instantiating the concept of an OAIS information package. A widely used example of the latter is the Metadata Encoding and Transmission Standard (METS), an XML-based document format that supports the encoding, or packaging, of the wide variety of metadata associated with archived digital objects. As McDonough (2006, 148) points out, ‘there was a desire for METS to facilitate the exchange and interoperability of digital library objects across digital library systems and to provide support for the long-term preservation of digital library objects by serving as a potential Submission Information Package, Archival Information Package and Dissemination Information Package within the [OAIS] Reference Model’.

METS is certainly not the only means of encoding OAIS archival information packages. For example, BagIt⁴⁴ is a file packaging specification jointly developed by the US Library of Congress and the California Digital Library. The Archivematica digital preservation system uses BagIt to create AIPs. Other encoding schema are also available, such as Fedora Object XML (FOXML)⁴⁵, although many of these are not explicitly linked to implementation of the OAIS AIP concept. Since a detailed description of all encoding schema would be more ambitious than space constraints allow, this section will focus on the METS standard, in view of its general association with the OAIS information package concept, as well as its popularity across a wide range of implementation contexts.⁴⁶

A METS document for a digital object includes a self-descriptive header; descriptive metadata about the object; administrative metadata (in particular, technical metadata, rights metadata, metadata about the analog source of the digital object, and provenance metadata); a list of the files constituting the digital object; a structural map of all the digital object’s components; a list of links that establish relationships between the components of the structural map; and a list of ‘behaviors’ that can be associated with the digital object. Metadata can be internally encoded within the METS document, or linked from an external source such as a registry.⁴⁷

A METS document serves as a structure for packaging together archived digital objects and their associated metadata. In this sense, it can be viewed as an implementation of the OAIS information package concept. Recall that an OAIS information package consists of the archived object, along with metadata necessary to support its long-term preservation, access, and understandability, bound into a single logical package. In the same way, a METS document combines a digital object and its metadata into a single logical package that can be ingested as a unit into the repository (i.e., as a Submission Information Package); managed as a unit within the repository system (i.e., as an Archival Information Package); or extracted and disseminated as a unit to parties external to the repository, such as the Designated Community or other repositories (i.e., as a Dissemination Information Package). In carrying these functions, a METS document acts as the packaging information defined in the OAIS information model: information that serves to combine, physically or

⁴⁴ <https://tools.ietf.org/html/draft-kunze-bagit-10>

⁴⁵ <http://fedora-commons.org/download/2.0/userdocs/digitalobjects/introFOXML.html>

⁴⁶ Errata: this paragraph contained a typographical error when originally published. The error was corrected on 09/12/2014

⁴⁷ See <http://www.loc.gov/standards/mets/METSOverview.v2.html> for a more detailed explanation of the sections of a METS document.

logically, all of the components of an information package, permitting them to be identified, located, and managed as a single logical unit.

Originally released in 2001, the METS standard is maintained by the US Library of Congress.⁴⁸ Given that a METS document provides packaging structure for the metadata associated with an archived digital object, and that the PREMIS Data Dictionary has emerged as a leading standard for implementing these metadata requirements (in particular, the metadata covered by the administrative metadata section of a METS document), the two standards are natural complements – and indeed, METS has become a popular vehicle for packaging and storing PREMIS-conformant metadata in digital repositories. In light of this, the PREMIS Editorial Committee has released a set of guidelines and recommendations for implementing PREMIS with METS.⁴⁹ A study by Vermaaten (2010) examines the use of PREMIS in registered METS profiles – that is, specific instantiations of a METS document by a particular institution or project – and proposes a checklist of implementation decision points to consider when using PREMIS with METS.⁵⁰

Another initiative aimed at practical application of the OAIS information package concept is the TIPR (Towards Interoperable Preservation Repositories) project.⁵¹ TIPR developed a common protocol for exchanging ‘rich archival information packages’ between heterogeneous repository systems. The Repository eXchange Package (RXP) is based on both PREMIS and METS, and supports the extraction of preservation metadata from a local repository system, packaging of the metadata into an exchange-ready format, and the transfer of the RXP-based AIP to another repository, where it can be easily ingested into the local repository system. Ensuring the feasibility of inter-repository transfers of archival information packages is an important aspect of long-term digital stewardship, and includes use cases such as ‘succession’, where a repository ends its operations and must transfer its archived content to another repository. In these circumstances, ‘Source repositories can package AIPs as RXPs and transfer the RXPs to successor repositories. The successor repositories can convert the RXPs into SIPs, ingest the SIPs, ingest and update the associated digital provenance, provide notification of receipt, and formally take custody of the content’ (Pawletko and Caplan, 2011, 2). As discussed above, the OAIS reference model itself notes a variety of incentives or rationales for promoting interoperability between OAIS-type archives. The TIPR protocol is a practical application of OAIS information package concepts that addresses OAIS interoperability by lowering the obstacles to the flow of OAIS information packages across repositories.

6.5. Other OAIS-related standards

The OAIS reference model provides fertile ground for a range of possible OAIS-related standards; indeed, the model’s architects anticipated from the beginning that OAIS would serve as a foundation for a variety of standard-building activities. The reference model suggests a number of areas where standards would be helpful, such as interfaces between archives; submission and ingest of digital objects into archives; delivery of digital objects from archives; metadata for digital objects; search and retrieval protocols; media and format migration; and certification of archives. The reference model notes several standards that have already emerged in some of these areas, including PREMIS (digital metadata) and TRAC (archive certification). A number of more general standards, such as a Data Entity Dictionary Specification Language for creating data product descriptions, or a suite of ISO Standards related to record management practices, are also noted (OAIS, 2012, 1-4 –1–5).

Several initiatives have focused on standardizing the interface between Producers and Archives. ISO 20652:2006 defines a Producer-Archive Interface – Methodology Abstract Standard (PAIMAS).⁵² PAIMAS

⁴⁸ <http://www.loc.gov/standards/mets/mets-home.html>

⁴⁹ <http://www.loc.gov/standards/premis/guidelines-premismets.pdf>

⁵⁰ <http://www.dlib.org/dlib/september10/vermaaten/09vermaaten.html>

⁵¹ <http://wiki.fcla.edu/TIPR>

⁵² See http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39577; also CCSDS 651.0-M-1 (Magenta Book), available at <http://public.ccsds.org/publications/archive/651x0m1.pdf>.

provides a standardized description of the interactions that take place between Producers and an OAIS-type archive. The standard segments the process of transferring information from the Producer to the OAIS into a set of distinct phases, and provides a detailed description of the anticipated outcome of each phase, as well as the set of actions which must take place to bring about this outcome. This framework is intended to serve as a basis for identifying areas within the Producer-Archive interface that would benefit from more focused standards, recommendations, and best practices, and also provides a foundation for the development of automated processes and software tools to support the information transfer process. PAIMAS also provides a more detailed exposition of the responsibilities and functions of the OAIS Ingest and Administration high-level services than what is provided in the OAIS reference model.

A new standard is currently under development that will provide ‘a standard method for formally defining the digital information objects to be transferred by an information Producer to an Archive and for effectively packaging these objects in the form of Submission Information Packages’ (PAIS, 2012, 1–1). The Producer–Archive Interface Standard (PAIS) is intended to support more precise definitions of the digital objects transferred from Producers to Archives, which in turn will help archives process and validate the objects received during the submission process. PAIS is being developed under the auspices of CCSDS, and in 2012 was granted Red Book status – i.e., the draft standard is considered technically mature and has been released for review by CCSDS member agencies. Taken together, PAIMAS and PAIS help shape the transfer of information from the Producer to an OAIS-type archive into a consistent, well-understood process, as well as cultivate a mutual understanding between Producers and archives in regard to their respective responsibilities and expectations as participants in the submission and ingest processes.

7. Conclusions

As the previous discussion suggests, the OAIS reference model has exerted a significant influence on the architectures, workflows, standards, and practice of digital preservation. Since its first publication in 1997, and later formalization as an ISO Standard in 2002, OAIS has become part of the bedrock supporting reliable long-term stewardship of digital materials, one of the fundamental documents that has guided digital preservation collaboration, development, and implementation over the last decade.

This *Technology Watch Report* has sought to re-introduce digital preservation practitioners to OAIS, by recounting its development and recognition as a standard; its key revisions; and the many channels through which its influence has been felt. It has now been more than a decade since the release of OAIS as an ISO Standard. This seems like a sufficient period from which to draw a reasonable assessment of the OAIS reference model's legacy – in other words, its most significant benefits to the digital preservation community, and on the other side of the ledger, some areas where the reference model perhaps fell short of expectations. In the case of the latter, these may also represent opportunities for future OAIS-related work.

7.1. The OAIS legacy so far

Perhaps the most important achievement of the OAIS reference model to date is that it has become almost universally accepted as the *lingua franca* of digital preservation. The concepts and terminology articulated in the reference model have become a useful shorthand for digital preservation practitioners; a means of shaping and sustaining conversations about digital preservation across disparate domains; and a general mapping of the landscape that stewards of our digital heritage must navigate in order to secure the long-term availability of the digital materials in their care. These three legacies of the model: a language, a shared reference point, a map – did much to consolidate knowledge of the problem space occupied by digital preservation at the time that this issue was attracting the notice of information managers across a wide range of domains. The OAIS reference model continues in this role today, and, if anything, has become even more entrenched as a foundation document in the field of digital preservation – evidenced by the frequent references (often without definition) to OAIS concepts and terminology in many papers and presentations.

In addition to providing a shared vocabulary for understanding, discussing, and supporting collaboration in digital preservation, the OAIS reference model also serves as something akin to a conceptual blueprint elucidating the fundamental components of a preservation repository, as well as the information objects the repository manages. As the discussion in the previous section illustrates, OAIS has been a useful starting point for many digital preservation initiatives that seek to operationalize or standardize one or more OAIS concepts. In a sense, these initiatives can be viewed as filling in the broad outlines of the map of the digital preservation landscape supplied by the reference model. Alignment with concepts defined in OAIS helps orient a technical implementation, draft standard, or other activity within the broader repository context that the OAIS reference model defines, making it part of a cohesive big picture.

One important marker for OAIS's success may be the fact that a comparison of the original (2002) version of the reference model with the current (2012) version reveals surprisingly little significant revision. Certainly some important elements have changed – for example, the addition of Access Rights Information to the PDI component of an AIP – but overall, the original version of OAIS has stood the test of time very well. In particular, the current versions of the OAIS Environment, Functional, and Information Models are essentially the same as those first promulgated more than a decade ago. Those responsible for the OAIS reference model in its original form can make a reasonable argument that for the most part, they got it right the first time.

The OAIS reference model has become something of a Platonic ideal for a digital preservation repository, the environment it operates in, and the information objects it manages. It conceptualizes, and provides a nomenclature for, the key features of these components, and illustrates how they collectively operate to meet the mandatory responsibilities of an OAIS-type archive. In doing so, the OAIS reference model supplies

a benchmark against which real-world repository implementations can be compared and evaluated, either in terms of how closely they approximate the ideal, or in what ways they depart from it. In assessing the impact of the OAIS reference model since its approval as an ISO Standard in 2002, it seems reasonable to conclude that OAIS has become a foundational resource for understanding digital preservation, a language for talking about digital preservation issues, and a starting point for implementing digital preservation solutions.

7.2. OAIS Impact: Some Limitations

While the influence of the OAIS reference model is difficult to question, it is possible to identify a few limitations associated with its impact, which correspondingly limits the overall value of OAIS to the digital preservation community. It is important to emphasize that these limitations are not intrinsic to the OAIS reference model itself, but rather to the nature of the work that builds upon OAIS, and how the reference model is perceived and applied.

As we have seen, many initiatives use OAIS concepts as a means of contextualizing their outputs. For example, the METS schema can be viewed as a possible implementation of the OAIS concept of an information package. However, METS is *not* a formal standard explicitly aimed at defining an implementation for an OAIS information package; in fact, an information package can be implemented in a variety of ways without use of the METS schema, and METS can be implemented without any reference to an OAIS information package. This example is emblematic of a broader problem with OAIS: very few of its concepts have been directly and formally operationalized as standards in their own right. The PAIMAS and PAIS initiatives (discussed in the previous section), which are aimed at standardizing the interface between Producers and Archives, are probably the purest (and arguably the only) examples of taking concepts and relationships directly from the OAIS reference model and defining them more precisely through a standardization process.

Given that more than a decade has passed since OAIS was itself approved as an ISO Standard, more progress perhaps might have been expected in developing formal standards that directly extend OAIS concepts into standardized forms. Of course, this is not an indictment of the reference model itself, and indeed the reference model itself takes pains to emphasize that it ‘does not specify a design or an implementation. Actual implementations may group or break out functionality differently’ (OAIS, 2012, 1–2). However, the reference model also notes repeatedly that a key rationale for its development was to promote standardization: for example, observing in one place that ‘Standards developers are expected to use this model as a basis for further standardization in this area. A large number of related standards are possible’ (OAIS, 2012, 1-2); and in another, that OAIS ‘should also provide a basis for more standardization and, therefore, a larger market that vendors can support in meeting archival requirements’ (OAIS, 2012, 1–3).

The discussion in the previous section shows that OAIS has indeed underpinned many important initiatives that have advanced the frontiers of digital preservation solutions. Yet the relationship between these initiatives and OAIS can be tenuous, if not obscure. A design, a protocol, even a standard can self-declare itself OAIS conformant (but without an explicit accounting of how conformance is actually manifested). Initiatives can use OAIS concepts as a means of labelling or describing various components within their structure (but these concepts can be used quite superficially, more as an expository shorthand rather than a detailed mapping); OAIS can be cited as a foundation or starting point for a particular initiative, or alternatively the initiative can declare itself informed by OAIS (but again, without necessarily any elaboration on how this was so). In short, OAIS may indeed be implicit or even ‘recoverable’ from these initiatives, but it is not necessarily explicit within, or intrinsic to, their design.

The rather indistinct path from OAIS concepts to OAIS-based implementations can lead to uncertainty over issues like OAIS conformance, which was discussed at some length in the previous section. Recall that this discussion concluded with the assessment that despite the risk of imprecision, OAIS conferred a variety of benefits as a unifying framework that supported discussion, influenced design, enhanced trust, and encouraged standards-building. However, it is useful to remember that an ‘OAIS-type archive’ is still one built

primarily on OAIS *concepts*, not an OAIS suite of *standards*. This then leads us to an important question: are there as yet unrealized benefits to be obtained from establishing a more concrete relationship between the OAIS reference model and real-world implementations through the direct, authoritative standardization of concepts embedded in the OAIS environment, functional, and information models?

This is a question that should be considered as part of a broader discussion of the future of the OAIS reference model. If we conclude that we do indeed lack sufficient standardization around OAIS concepts, then this is probably the best indicator of where priorities for future OAIS-related work should lie. There is still a great deal of scope for standardization around the concepts and relationships detailed in the OAIS reference model. The digital preservation community would benefit from a careful assessment of where more precise and authoritative definitions of those concepts and relationships would accelerate progress in achieving robust, widely applicable, and interoperable digital preservation solutions.

7.3. OAIS: a theory of digital preservation

In a paper exploring the application of OAIS concepts to the preservation of computer games, McDonough observes that the ‘OAIS reference model provides a solid theoretical basis for digital preservation efforts ... As seen in our efforts to attempt to apply concepts from the OAIS reference model in the packaging of computer games for long-term preservation, however, theory and practice can sometimes have an uneasy fit.’ (McDonough, 2012, 1631). This is probably as good a way as any to sum up both the impact and the limitations of the OAIS reference model. OAIS supplies a conceptual – a theoretical – view of the functional components of a digital repository, the environment in which it operates, and the information objects that it manages. It is a useful framework for thinking about the salient features of a generalized process of digital preservation, with the detail of implementation choices and context abstracted away – much like a theory of any natural or social phenomenon necessarily strips out the idiosyncrasies of the real-world manifestations they purport to represent. As such, there is always a gap between theory and practice. But the worth of most theories is bound up in their ability to illuminate the features of a particular space, and to inform decision-making around real-world activities that take place in that space. Measured against this yardstick, the OAIS reference model, as a theory of digital preservation, has performed with distinction.

8. Glossary of Abbreviations

AIC	Archival Information Collection
AIP	Archival Information Package
AIU	Archival Information Unit
ASCII	American Standard Code for Information Interchange
CCSDS	Consultative Committee for Space Data Systems
DAITSS	Dark Archive in the Sunshine State
DIP	Dissemination Information Package
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
GIF	Graphic Interchange Format
HTML	Hypertext Markup Language
ICPSR	Inter-university Consortium for Political and Social Research
ISO	International Organization for Standardization
JPEG	Joint Photographic Experts Group
LOCKSS	Lots of Copies Keeps Stuff Safe
METS	Metadata Encoding and Transmission Standard
NARA	US National Archives and Records Administration
OAIS	Open Archival Information System
OCLC	Online Computer Library Center, Inc.
PAIMAS	Producer-Archive Interface – Methodology Abstract Standard
PAIS	Producer-Archive Interface Standard
PDF	Portable Document Format
PDI	Preservation Description Information
PREMIS	Preservation Metadata: Implementation Strategies
RLG	Research Libraries Group
RXP	Repository Exchange Package
SGML	Standard Generalized Markup Language
SIP	Submission Information Package
TIFF	Tagged Image File Format
TIPR	Towards Interoperable Preservation Repositories
TRAC	Trustworthy Repositories Audit & Certification: Criteria & Checklist
XML	Extensible Markup Language

9. References

Center for Research Libraries and OCLC 2007, *Trustworthy repositories audit and certification: Criteria and checklist* (Version 1.0), The Center for Research Libraries and Online Computer Library, Inc., Chicago IL and Dublin OH. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (last accessed 01/08/14)

Consultative Committee for Space Data Systems Secretariat 2012, *Producer–archive interface specification (PAIS): Draft recommended standard* (CCSDS 651.1-R-1: Red Book), CCSDS, Washington, DC. <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206511R1/Attachments/651x1r1.pdf> (last accessed 01/08/14)

Consultative Committee for Space Data Systems Secretariat 2012, Reference model for an open archival information system (OAIS): Recommended practice (CCSDS 650.0-M-2: Magenta Book), CCSDS, Washington, DC. <http://public.ccsds.org/publications/archive/650x0m2.pdf> (last accessed 01/08/14)

Lavoie, B. and Gartner, R. 2013, *Preservation Metadata* (2nd edition), Digital Preservation Coalition, York. <http://dx.doi.org/10.7207/twr13-03> (last accessed 01/08/14)

McDonough, J. 2006, 'METS: Standardized encoding for digital library objects', *International Journal on Digital Libraries* 6, 148–158.

McDonough, J. 2008, 'Structural metadata and the social limitation of Interoperability: A sociotechnical view of XML and digital library standards development', Balisage: The Markup Conference 2008. <http://www.balisage.net/Proceedings/vol1/print/McDonough01/BalisageVol1-McDonough01.html> (last accessed 01/08/14)

McDonough, J. 2012, "'Knee-deep in the data'": Practical problems in applying the OAIS reference model to the preservation of computer games', Hawaii International Conference on System Sciences 2012. http://www.hicss.hawaii.edu/hicss_45/bp45/dm1.pdf (last accessed 01/08/14)

Pawletko, J. and Caplan, P. 2011, 'Towards interoperable preservation repositories: Repository exchange package use cases and best practices', IS&T Archiving Conference 2011. https://fclaweb.fcla.edu/uploads/is_and_t-pawletko-caplan-final.pdf (last accessed 01/08/14)

Vermaaten, S. 2010, 'A checklist and a case for documenting PREMIS-METS decisions in a METS profile', *D-Lib Magazine* 16, online at: <http://www.dlib.org/dlib/september10/vermaaten/09vermaaten.html> (last accessed 01/08/14)

Waters, D. and Garrett, J. 1996, *Preserving Digital Information: Report of the task force on archiving of digital information*, Council on Library and Information Resources, Washington, DC. <http://www.clir.org/pubs/reports/reports/pub63watersgarrett.pdf> (last accessed 10/09/14)