OAIS VERSION 3 DRAFT UPDATES

David Giaretta
PTAB Ltd
UK
david@giaretta.org
0000-0001-8414-7509

John Garrett
Garrett Software
USA
garrett@his.com
0000-0003-4460-9902

Mark Conrad
NARA
USA
mark.conrad@nara.gov

Eld Zierau
Royal Danish Library
Denmark
elzi@kb.dk
0000-0003-3406-3555

Terry Longstreth
Washington Acad. Sciences
USA
longstreth@acm.org
0000-0001-8228-6185

J Steven Hughes
Jet Propulsion Laboratory
USA
john.s.hughes@jpl.nasa.gov
0000-0003-4851-293X

Matthias Hemmje
FernUniversität Hagen
Germany
matthias.hemmje@fernuni-hagen.de
0000-0001-8293-2802

Felix Engel
FernUniversität Hagen
Germany
felix.engel@fernuni-hagen.de
0000-0002-3060-7052

Abstract – This paper provides a high-level view of the changes proposed in the OAIS Reference Model.

Keywords – OAIS, digital preservation

Conference Topics – Designing and Delivering Sustainable Digital Preservation; Exploring New Horizons.

I. INTRODUCTION

The aim of this paper is to describe and explain the most significant updates which have been made to version 2 OAIS [1], which was published in 2012, from the point of view of the authors, who have all been deeply involved with the revision. These updates resulted in a draft which, at the time of writing, is the text to be submitted for the Consultative Committee for Space Data Systems (CCSDS) and ISO review. Further changes may be made, either before submission or during the review process, after which version 3 of OAIS will be published.

In the body of this paper the text in italics is taken from the current draft.

ISO 16363, which is the basis of auditing trustworthy digital repositories, is being updated to be consistent with the changes to OAIS.

II. OAIS REVIEW PROCESS

The previous update of OAIS was primarily debated at physical CCSDS meetings augmented by email exchanges via a mailing list, together with weekly telecons. This did allow wide participation but the CCSDS MOIMS-DAI [2] working group, which oversees the latest revision, wanted to improve and widen access. With this aim in mind, the http://review.oais.info website was set up, which allowed anyone to register and contribute to the discussion. It allowed everyone, whether registered or not, to view all the contributions and discussions.

The page for each suggested change showed the various contributions and the final decision on whether to reject, accept or modify that proposal. Changes to the evolving document were reviewed and dispositioned at weekly telecons and by in-

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

DOI: 10.1145/nnnnnnn.nnnnnnn

iPRES 2019

person and remote attendees of MOIMS/DAI sessions at the semi-annual CCSDS plenary meetings. The disposition reflected the consensus reached in these telecons and meetings, as reflected by agreements recorded in http://review.oais.info.

A marked-up Word document was maintained, with comments linking each change to the discussion on the website.

When all 200+ suggested changes to OAIS had been resolved, a second round of comments were collected on this marked up document to check for inconsistencies and small errors. These were resolved via the telecons and appropriate changes made, with comments to record the justification for the change.

The remainder of this document describes the major changes made in the draft which will go for formal CCSDS and ISO review.

To some the changes described here may seem unimportant or irrelevant but they have been made in anticipation of new challenges to the preservation community which may arise over the 5 or more years before version 4 will be produced.

III. UPDATES TO OAIS CONCEPTS

A. Representation Information

One of the key OAIS concepts is Representation Information, which, when combined with a Data Object, produces an Information Object. The question as to how much Representation Information is needed is determined by the definition of the Designated Community and its Knowledge Base.

The amount of Representation Information will change over time as the Knowledge Base of the Designated Community changes. The OAIS needs to ensure that it has Long Term access to all the relevant Representation Information. A choice must be made whether the OAIS collects all the relevant Representation Information itself or references the existence of the Representation Information in another trusted OAIS Archive. That choice is an implementation and organization decision.

The updates make it clear that **in special cases** the initial amount of Representation Information required may be very minimal. For example, *for a specific Data Object and a specific Designated Community, the Knowledge Base of the Designated Community is adequate for its members to understand or use the Data Object. In such cases the Representation Information could be the statement that no additional Representation Information is needed for that specific Designated Community at that particular time*, but further Representation Information may need to be collected in future. The revised text goes on to say, …"*any Representation Information that can be gathered at ingest should be included since it will likely be costlier to rediscover and add it at a later time*."

B. Preservation Description Information (PDI)

In the versions of OAIS up to now the components of PDI, namely Provenance Information, Reference Information, Fixity Information, Access Rights Information and Context Information, referred to the Content Information, i.e. the Content Data Object plus its Representation Information. Although these are a consistent and useful set of definitions, it does cause some problems in terms of potential implementations. Consider the case where one deals with a distributed network of Representation Information, which changes with the Designated Community's Knowledge Base. A change in some part of the Representation Information network would mean that all the elements of the PDI would change.

The update concerning PDI is that all the components of PDI would now refer to the Content Data Object rather than Content Information.

There are several reasons for this change. The consensus was that for most, perhaps all, repositories, the PDI components do refer to the Content Data Objects. For example, the Fixity Information is often essentially a digital digest of the Content Data Object. This focus on Data Objects would also make audits of repositories more practical since the auditor can perform checks on specific Content Data Objects. Of course, even the Content Data Object may be complex, for example consisting of many files, but at least changes in the Knowledge Base of the Designated Community does not cause it to change.

A related point considered by the group was that, for example, the Representation Information should have Fixity also. To clarify this point the following note was added to emphasize the fact that, from the very first version of OAIS, the Information Model applies to every one of the things which are called "Information", including, for example, Representation Information and Provenance Information.

Defining PDI (as well as its components - Provenance Information, Context Information, Reference Information, Fixity Information, and Access Rights Information) as relevant to the Content Data Object does not mean that those concerns are any less important for other data objects or at other levels, for example, it is important to apply reference, fixity, provenance, context and access rights to Representation Information, or to any other information the Archive is preserving. Definition of these terms as relevant to the Content Data Object is simply to ease discussion of these concepts at the Content Data Object level.

In other words when one is talking about, for example, Representation Information as the target of preservation, then one can regard it as Content Information in its own right, as well as being part of another instance of Content Information. To some readers this may seem a strange way to describe things, but it is similar to what should be the familiar arrow in the OAIS Information Model which "loops back" from Representation Information back to itself.

C. *Preservation Objectives*

Usability has played a central role in defining preservation. However, there was a feeling that the meaning of usability needed to be clearer, and more testable. To this end the concept of a "Preservation Objective" has been introduced and defined as a *specific achievable aim which can be carried out using the Information Object*.

Preservation Objectives can then be used in the definition of other terms including:

- *Representation Information: The information that maps a Data Object into more meaningful concepts. so that the Data Object may be understood in ways exemplified by Preservation Objectives*.
- *Independently Understandable: A characteristic of information that is sufficiently complete to allow it to be understood by the Designated Community, as exemplified by the associated Preservation Objectives, without having to resort to special resources not widely available, including named individuals*

Preservation Objectives are intended to allow the repository to make it possible to test and demonstrate whether the information actually is Independently Understandable by members of the Designated Community now and into the future.

Examples of Preservation Objectives are provided in the updated OAIS:

- *The ability to render documents, images, videos or sounds in a way which is sufficiently similar to the original. This could be checked by verifying that, for example, the document is readable or the image is viewable. An analysis of the colours could also be compared. A spectral analysis could be performed on the sounds and compared with that of the original.*
- *The ability to process a dataset and generate the data products expected. This could be checked by comparing with something generated earlier, for example on Ingest.*
- *The ability to understand a dataset and use it in analysis tools to generate results, for example the density of electrons in the upper atmosphere or the structure of a molecule, given certain measurements. These could be compared with results generated earlier.*
- *The ability to re-perform an artistic performance. This could be compared with a recording of a previous performance.*

IV. UPDATE TO THE OAIS FUNCTIONAL MODEL

There have been many small clarifications made to the various text and diagrams which make up the Functional Model, introducing unambiguous shapes for diagram entities; MOIMS-DAI hopes that CCSDS/ISO will allow the publication of the new version to include the colors which give visual clues as to the grouping of the boxes.

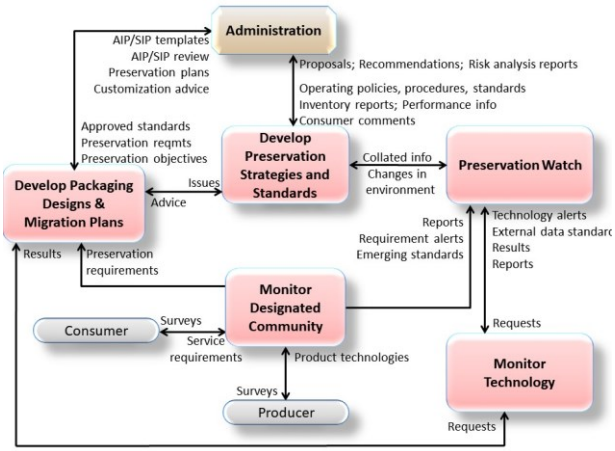

Figure IV-1~~IV-1~~ Updated Preservation Planning Functional Entity

The one area where an extra function has been added is to the Preservation Planning Functional Entity.

The additional function is the already widely used “Preservation Watch”. This is described in the update as follows:

The Preservation Watch function is the role of collating preservation related information from a variety of internal and external entities. The Preservation Watch function also brings in reports, requirement alerts and emerging standards from the Monitor Designated Community function and technology alerts, external data standards, results and reports from the Monitor Technology function. Changes in the environment of the Archive (financial, political, and environmental) can be part of the Preservation Watch function.

Previously, Preservation Watch functionality was primarily located within the Develop Preservation Strategies and Standards.

V. UPDATES TO THE OAIS INFORMATION MODEL

The major updates to the Information Model carry forward the changes which have been described in section III. These are summarized in the following diagram where the PDI connects to the Data Object rather than the Content Information:

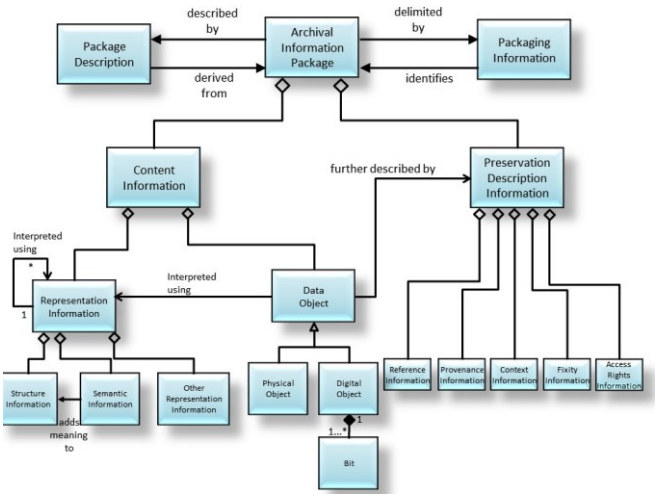

Figure V-~~1~~V-~~1~~ Updated AIP diagram

A. Updates to Information Package Definition

An Archival Information Package is the most detailed example of an Information Package, one which must contain Content Information as well as PDI. However, SIPs and DIPs do not need to contain any of these components.

The update to the general OAIS Information Package, shown in Figure V-2.

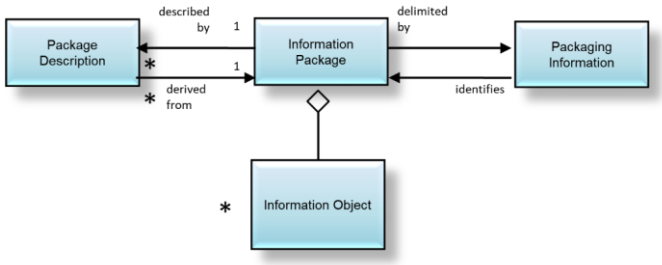

Figure V-2 Updated Information Package

This change makes it clear that SIPs and DIPs can be defined in a much more flexible way. Note that this does not require any changes to the definition of the AIP because, as illustrated in Figure V-3, the combination of Content Information and PDI can be regarded as a single, albeit complex, Object, made up of multiple Information Objects.

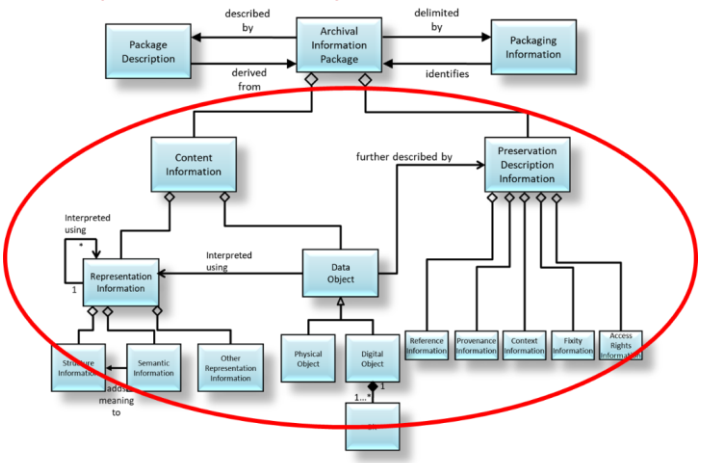

Figure V-3 The combined Information Object in an AIP

VI. UPDATES TO PRESERVATION PERSPECTIVES

Major changes have been made to the section of OAIS which describes practices that have been, or might be, used to preserve digital information and to preserve access services to digital information.

Up to now, essentially the only preservation practice which has been explicitly described has been Migration and Preservation of Access, e.g. Emulation. However clearly the OAIS mandatory responsibilities require that there be adequate Representation Information, and that the OAIS should preserve information against all reasonable contingencies, including the demise of the Archive.

The changes in the new draft now include explicitly that the Content Data Object being preserved may be

(1) kept by the Archive but may be changed or
(2) kept by the Archive unchanged or

(3) not kept by the Archive, but instead be handed on to another Archive

Each of these three imply the following:

In case (1) the Archive may Transform the Content Data Object

In case (2) the Archive may add Representation Information to ensure the Content Information is Independently Understandable

In case (3) the Archive may hand over the AIP which contains the Content Data Object

This change makes the text as a whole more consistent and clearer.

VII. UPDATES TO ARCHIVE INTEROPERABILITY

A major change to the discussion of various possible types of archive interactions is the way in which the distribution of OAIS functionality may be described. Such a distribution of functionality could be motivated, for example, by cost reduction or the availability of a comprehensive functionality offer. These descriptions should allow archives to be described more accurately and make it even clearer that an OAIS has never been required to be a monolithic organisation.

The text describes some possible categories (not an exhaustive or mutually exclusive list) of Archive associations. The first set of three categories has successively higher degrees of organizational interaction:

- *Independent: Archives motivated by local concerns with no management or technical interaction among them.*
- *Cooperating: Archives with potential common Producers, common submission standards, and common dissemination standards, but no common Finding Aids.*
- *Federated: Archives with both a Local Community (i.e., the original Designated Community served by the Archive) and a Global Community (i.e., an extended Designated Community) which has interests in the holdings of several OAIS Archives and has influenced those Archives to provide access to their holdings via one or more common Finding Aids*.

Another set of categories, somewhat orthogonal to the previous set, differentiates according to how internal Archive functions and functional areas are addressed and by styles of resource sharing.

- *All In-house: Archives that perform all archival functions in-house.*
- *Shared resources: Archives that have entered into agreements with other organizations to share resources, perhaps to reduce cost. This requires various standards internal to the Archive (such as ingest-storage and access-storage interface standards) but does not alter the user community's view of the Archive.*
- *Distributed: Archives that have distributed the OAIS functionality either geographically or organizationally. Different levels, forms and organization of the distribution are possible. In every case, the Archive is required to oversee and manage the Archive's use of the distributed functions, but does not alter the user community's view of the Archive*

An important classification of distribution is where the supporting organizations, which supply the required functionality, are themselves each an OAIS. One can describe the arrangement as a primary OAIS using one or more supporting OAIS for specific services. In such a case, each supporting OAIS, as well as the primary OAIS must fulfill all requirements for OAIS conformance, namely the Mandatory Responsibilities and support for the Information Model. Therefore, service level agreements are required to guarantee proper implementation of the functionality distribution. Particularly, the primary OAIS must monitor that the supporting OAIS is meeting its service agreement. The conformance of each supporting OAIS may be used as a piece of evidence.

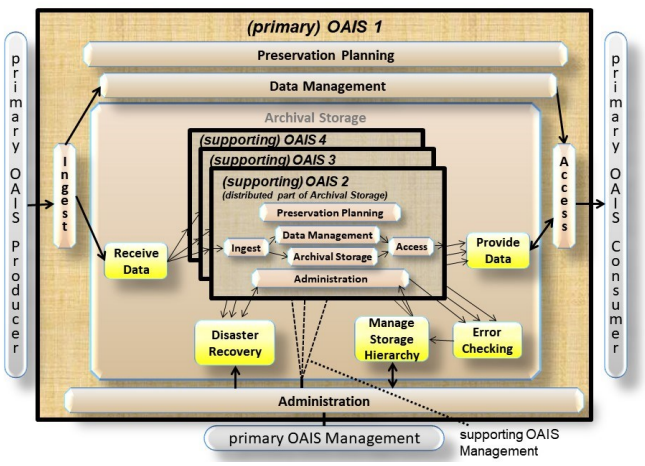

Figure VII-1~~VII-1~~ Primary/Supporting OAIS distributed functionality

The term 'Outer OAIS-Inner OAIS' has been used in the literature [3]. This usage is consistent with the "Outer OAIS" being the primary OAIS and the "Inner OAIS" being the supporting OAIS in cases where the "Outer OAIS" and "Inner OAIS" are each totally

conformant to OAIS requirements. To exemplify the use of distributed functionality with supporting (inner) OAISes the Figure VII-1 shows how a set of supporting OAISes complete the functionality of the primary OAIS Archival Storage.

VIII. CONCLUSIONS

The updates made to the current version of OAIS, to be submitted for CCSDS and ISO review, provide significant clarifications and, when integrated into ISO 16363, improve the auditability of repositories, for example by giving auditors specific tests to verify understandability by using the Preservation Objectives, where they are available.

The changes add further clarity to OAIS and bring in a number of useful concepts developed by others since version 2 of OAIS was published. They will allow repositories to be described more clearly, despite increasing complexity. The consensus was that the updates will not require archives which are currently conformant to OAIS to make major changes but will instead allow such archives to provide evidence about their conformance more clearly. In addition, the changes should keep OAIS fit for purpose as archives are faced with new challenges in the coming years.

ACKNOWLEDGEMENTS

Thanks are due to all the participants in the CCSDS MOIMS-DAI working group and others, in particular those who contributed over 200 suggested changes to OAIS and helped in their resolution.

REFERENCES

[1] Reference Model for an Open Archival Information System (OAIS). Magenta Book. Issue 2. June 2012, available from https://public.ccsds.org/Pubs/650x0m2.pdf also known as ISO 14721:2012.

[2] Data Archive Interoperability Working Group https://cwe.ccsds.org/moims/default.aspx#_MOIMS-DAI

[3] "Supporting Analysis and Audit of Collaborative OAIS's by use of an Outer OAIS – Inner OAIS (OO-IO) Model" by Eld Zierau and Nancy McGovern. In Proceedings of the 11th International Conference on Preservation of Digital Objects (iPres) 2014, pp. 209-218, available at http://www.ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf