OAIS 5-year review
InterPARES 2 and OAIS

October 30, 2006

Submitted on behalf of InterPARES 2 by:
Mahnaz Ghaznavi (Getty Research Institute, J. Paul Getty Trust), Sally Hubbard (Getty Research Institute, J. Paul Getty Trust), Jim Suderman (City of Toronto), Luciana Duranti (University of British Columbia), and Randy Preston (University of British Columbia)

INTRODUCTION

On the occasion of the fifth anniversary of its publication, the International Standards Organization (ISO) and Consultative Committee for Space Data Systems (CCSDS) ask whether changes are required to the existing reference model for open archival information system (OAIS), *ISO 14721:2003 Space data and information transfer systems -- Open archival information system -- Reference model*. Specifically, ISO and CCSDS are soliciting recommendations for any modifications to "reduce ambiguities or improve missing or weak concepts," and assistance with "the identification of outdated material."[1]

The findings of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project suggest some modifications to *ISO 14721:2003* are indeed warranted. The recommendations that follow draw on InterPARES research, which examines the requirements for and impediments to the creation, maintenance, and long term preservation of authentic, reliable, accurate records generated in complex digital environments. While the focus of OAIS is "information" rather than "records," we believe that some InterPARES findings are relevant to the review of the existing reference model, for the reasons outlined below.

---

[1] The "Reference Model for an Open Archival Information System (OAIS)" was approved in January 2002 as CCSDS 650.0-B-1, as ISO 14721 in 2003; the quotation is drawn from the call for comments announcement of the open comment period in advance of its five year review, http://nssdc.gsfc.nasa.gov/nost/isoas/oais-rm-review.html.

UPDATES TO ADD MISSING CONCEPTS OR STRENGTHEN WEAK CONCEPTS

Archives and authenticity

CCSDS characterizes OAIS as:

> an archive, consisting of an organization of people and systems, that has accepted the responsibility to **preserve information** and make it available for a Designated Community.[2]

The reference model goes on to propose:

> [t]he term "archive" has come to be used to refer to a wide variety of storage and preservation functions and systems. Traditional **archives** are understood as facilities or organizations which preserve records, originally generated by or for a government organization, institution, or corporation, for access by public or private communities. The **archive** accomplishes this task by taking ownership of the records, ensuring that they are understandable to the accessing community, and managing them so as to **preserve** their **information content** and **authenticity**.[3]

By contrast, OAIS defines "archive" as:

> [a]n organization that intends to preserve information for access and use by a Designated Community.[4]

To provide the broadest possible application, the authors of the reference model move away from incorporating into OAIS an inherited definition of "archive" that confines the activity to the preservation of merely "records"—a definition the authors anchor in the discourses of traditional archives and "archival science."[5] Within these discourses, however, there are marked differences between "archive" and "archives." The term "archive" is strictly British and refers to one "fonds," that is "[t]he whole of the records that a physical or juridical person accumulates by reason of its function or activity."[6] It should not be part of the

---

[2] Emphasis added; CCSDS, 1-1

[3] Emphases added; CCSDS, 2-1.

[4] CCSDS, 1-8.

[5] CCSDS, 1-7.

[6] InterPARES 2 Terminology Database, http://www.interpares.org/ip2/ip2_terminology_db2.cfm; see also the entry for "archive" in the Society of American Archivists Glossary, which notes that "United States and Canadian archivists generally deprecate the use of 'archive' (without an s) as

terminology of any model that refers to archives, which in North American English is the only term used for one or more fonds, that is, for archival material in general, *and* for the organization or program preserving it.

In the place of "records," the OAIS definition uses the broader term "information" to signify the object of preservation, and maps the concept of "records" to the reference model's Archival Information Package's (AIP) Content Information.[7] While the traditional definition of archives highlights authenticity, the current OAIS definition of archive does not explicitly remark upon it. Researchers involved in the second phase of InterPARES (IP2), which began in 2002, have been examining issues of authenticity, reliability and accuracy in the preservation of digital records, focusing on records produced in complex digital environments in the course of artistic, scientific and e-government activities. IP2 has used the following terminology to describe what is to be preserved:

> Data: *n.,* The smallest meaningful units of information.
> Document: *n.,* An indivisible unit of information constituted by a message affixed to a medium (recorded) in a stable syntactic manner. A document has fixed form and stable content.
> Record: *n.,* A document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference.[8]

IP2 builds on the work of the first phase of IP research. A key finding of the first phase of IP, conducted between 1999 and 2001, is that it is not possible to preserve electronic records but only the ability to reproduce them.[9] Put this way, the question is not whether the materials being preserved are records or information, per se. Instead the challenge has to do with the process of preservation and the presumption of authenticity any archives is able (and,

---

a noun to mean a collection of records ('archives'), but that form is common in other English-speaking countries. In information technology, the s-less form, 'archive', is commonly used as a verb and to describe collections of backup data'," http://www.archivists.org/glossary/index.asp.
[7] CCSDS, 1-7.
[8] InterPARES 2 Terminology Database.
[9] Ken Thibodeau, "Certifying Authenticity of Electronic Records: Interim Report of the Chair of the Preservation Task Force to the InterPARES International Team," unpublished report, 19 April 2000, http://www.interpares.org/documents/hm_saa_2000.pdf.

arguably required) to make when ingesting information packages for storage and/or dissemination. Regardless of whether the packages in question are records, their preservation as digital objects, and the designated communities' abilities to reuse the objects, rely on some measure of authenticity.[10]

In the place of "records" OAIS uses the concept of the Archival Information Package's (AIP) Content Information.[11]   According to the reference model, Preservation Description Information (PDI) is predicated on a clear definition of Content Information; the two together form the AIP.[12]   The idea of the content information being authentic is implicit here.

InterPARES research promulgates the notion that "an authentic record is a record that is what it purports to be and is free from tampering or corruption."[13] While the OAIS definition refrains from any explicit mention of authenticity, it does offer a definition of fixity as it:

> provides a wrapper, or protective shield, that protects the Content Information from undocumented alteration. For example, it may involve a check sum over the Content Information of a digital Information Package.[14]

However, what the OAIS model defines as fixity has no relationship with authenticity in the archival or even legal sense. Fixity is a technical benchmark that allows the preserver to authenticate what is stored (the AIP) against what was received (the SIP), and possibly against what is disseminated (the DIP). In this sense, fixity is but one characteristic among many for helping verify authenticity; however, it is important to clarify that it is not synonymous with the

---

[10] The observation that authenticity is an issue for the preservation not only of records, but also of other digital objects can also be found in Seamus Ross and Margaret Hedstrom, "Preservation research and sustainable digital libraries," *International Journal on Digital Libraries* 5, no. 4 (2005): 317-324, http://www.dljournal.org/.
[11] See note 6, above.
[12] "Only after the Content Information has been clearly defined can an assessment of the Preservation Description Information be made," CCSDS, 2-6.
[13] InterPARES Authenticity Task Force Report, http://www.interpares.org/book/interpares_book_d_part1.pdf
[14] CCSDS, 2-6.

concept of authenticity, nor does it provide a mechanism for the preserver to attest to the authenticity of the source records or information, because it does not allow any comparison of what is stored (the AIP) against what was created and used before ingest (pre-SIP).

The OAIS model insufficiently addresses the relationship between the creator (the producer in OAIS terms) and the preserver from the point of the creation of information. The reference model could extend guidelines for Producer Interaction (2.3.2), the negotiation for and acceptance of information (3.2.1) and the functions of ingest (4.1.1.2) to allow or require that Content Information (4.2.1.4.1) include an assertion of authenticity. Alternatively, if the OAIS model assumes the authenticity of submitted information (and one of the findings suggested by the IP2 research was that the creator be considered the arbiter of authenticity as long as information was in the creator's custody), or makes no judgment upon it, this should be explicitly stated.

OAIS further associates "fixity" with "authentication" when it defines "fixity information" as:

> The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information object has not been altered in an undocumented manner.[15]

InterPARES findings emphasize that, while there is a relationship between the two, the distinctions between authenticity and authentication are important ones to maintain in the discussion of digital preservation.[16]

> *Authenticity* [is a quality of the record and means] that [it] is not changed or manipulated after it has been created...or migrated over the whole continuum of ...creation, maintenanceand preservation...There is no relative degree of authenticity, while there may be for reliability. The status of being authentic, however, can change at any moment as a result of residual effects of an action or migration that has been performed on the

---

[15] CCSDS, 1-10.
[16] InterPARES Authenticity Task Force Report.

record over time. This is the case for digital data as well. By contrast, authentication is the process of guaranteeing the authenticity of a record. If authenticity is the status of being authentic, then authentication is the action or set of activities that demonstrate that something is authentic.[17]

While "reliability" is a quality that is established when digital objects are made, "accuracy" is a quality that exists (or not) when an object is generated, and the latter is at risk throughout the object's life-cycle. "Authenticity" is a quality that arises when an object is set aside for maintenance and preservation. It is a quality that must be protected through the on-going maintenance and preservation of the object. Authenticity is then a quality that is presumed by the archives and must be not only negotiated during ingest, but also maintained in the preservation system.

While OAIS is flexible enough to build in features to account for authenticity, it would benefit users of the reference model to have clarification and refinement of its role in the archives of today and tomorrow.

Recommendation: Disambiguate **"authenticity"** and **"authentication"** in the OAIS reference model and discuss the role of each in the preservation process.

Recommendation: Disambiguate **"authenticity"** and **"fixity"** in OAIS reference model and discuss the role of each in the preservation process.

Recommendation: Consider how authenticity relates to Ingest and Community Standards.

Representation Networks

The OAIS model states that Representation Information, made up of Structure Information and the Semantic Information, may contain references to other Representation Information, and thus form Representation Networks that may be extremely complex. However, it is unclear whether the model sufficiently

---

[17] Shelby Sanett and Eun Park, "Authenticity as a Requirement of Preserving Digital Data and Records," *IASSIT Quarterly* 24, no.1 (2000): 15-18, http://iassistdata.org/publications/iq/iq24/iqvol241sanett.pdf.

acknowledges the dependencies that an information object may be subject to (dependencies are not explicitly discussed outside section 5.2.2.2 on emulation). For example, the IP2 research on experiential systems found that "documents that enable performance or production of a work" may include "both the documents which describe the work and/or the instruments, devices or other things used in the performance of the work, and those which provide instructions on how to perform the work".[18]

Recommendation: Clarify whether the definition of Representation Networks more explicitly requires the documentation of all information that may be required to reproduce an information object; for instance, performance directions, or software patches. In other words, does content information as presently defined by OAIS provide sufficiently nuanced Representation Information that, when combined with PDI, produces a valid AIP for the type of objects that experiential, dynamic and interactive systems generate?

IDENTIFICATION OF OUTDATED MATERIAL

The following sentence from the 2002 reference model identifies something for release in 1999.

> The archive system is currently being validated. The service is planned to be made available to the scientific community on September 1999.[19]

Recommendation: Revise to update, or otherwise remove sentence.

---

[18] Luciana Duranti and Kenneth Thibodeau, "The Concept of the Record in Interactive, Experiential, and Dynamic Environments: the view of InterPARES," *Archival Science* 6, no1 (2006): 13-68, http://www.dljournal.org/.
[19] CCSDS, A-22.