

 [COVID-19 | Information and resources to help \(https://www.oclc.org/en/covid-19.html\)](https://www.oclc.org/en/covid-19.html)

(<https://www.oclc.org/research.html>)



Meeting the challenges of digital preservation: The OAIS reference model

by Brian Lavoie

Digital information offers both opportunities and challenges for libraries in their traditional role as custodians of society's accumulated knowledge. While embodying clear advantages--precise replication, machine processing, online content delivery--information in digital form also introduces a host of difficulties with regard to access and preservation. The life span of digital storage media can be surprisingly short, and the rapid evolution of rendering technologies can impede future access. The advantages and disadvantages of digital information have been well documented in numerous sources.

What is required to preserve and maintain access to digital information over the long term? This question is still far from being satisfactorily answered, but a recent initiative by NASA's Consultative Committee for Space Data Systems offers common ground for discussion. The Open Archival Information System (OAIS) reference model is a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term. The purpose of the reference model is to increase awareness and understanding of concepts relevant for archiving digital objects, especially among nonarchival institutions; elucidate terminology and concepts for describing and comparing data models and archival architectures; expand consensus on the elements and processes endemic to digital information preservation and access; and create a framework to guide the identification and development of standards.

Although the OAIS is sufficiently general to encompass archives of physical as well as digital objects, it is in the context of the latter that the OAIS obtains its impetus. The reference model has been well-received by a diverse community of institutions interested in the long-term preservation of digital information. A number of digital initiatives in the library community, such as the CEDARS, PANDORA and NEDLIB projects, have either adopted the OAIS model as the conceptual framework behind their digital preservation efforts, or have been informed by its conclusions.

The OAIS reference model is currently a draft International Standardization Organization (ISO) standard and is expected to become a full-fledged standard in the future. As such, it is likely that the OAIS will be a highly visible component of the ongoing effort to address the challenges of preserving digital information.

Background

The Consultative Committee for Space Data Systems (CCSDS) was established in 1982 to provide an international forum for space agencies interested in the collaborative development of standards for data handling in support of space research. In 1990, CCSDS entered into a cooperative agreement with Subcommittee 13 (space data and information transfer systems) under Technical Committee 20 (aircraft and space vehicles) of the ISO, whereby CCSDS recommendations would undergo normal ISO review and voting and, eventually, evolve into ISO standards.

At the request of the ISO, CCSDS assumed the task of coordinating the development of archive standards for the long-term storage of digital data. To initiate this process, a reference model was developed to establish common terms and concepts, provide a framework for elucidating the significant entities and relationships among entities in an archive environment, and serve as the foundation for the development of standards supporting the archive environment. CCSDS's efforts resulted in the release of the OAIS reference model draft recommendation in May 1999.

Open Archival Information System

An OAIS is understood to mean any organization or system charged with the task of preserving information over the long term and making it accessible to a specified class of users (known as the Designated Community). The use of the word "open" in OAIS refers to the fact that the model and future recommendations associated with the model are developed in open forums; it does not make any presuppositions concerning the level of accessibility of information in the archive.

An OAIS-type archive is expected to meet certain minimum responsibilities:

- negotiate and accept appropriate information from information producers

- obtain sufficient control of the information to ensure long-term preservation
- determine the scope of the Designated Community
- ensure the information is understandable by the Designated Community without the assistance of the information producers
- follow documented policies and procedures to ensure the information is preserved against reasonable contingencies, and to enable the information to be disseminated as authenticated copies of the original or as traceable to the original
- make the information available to the Designated Community

The OAIS reference model details a conceptual design for an archive, including its primary components and their associated functions and relationships, to support these requirements.

The OAIS Environment

The reference model's specification of the environment for an OAIS-type archive is shown in Figure 1.

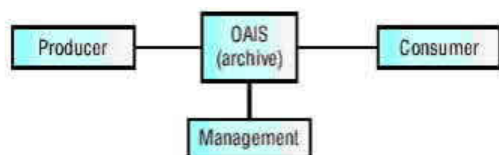


Figure 1

The OAIS environment is derived from the interaction of four entities: producers, consumers, management and the archive itself. Producers supply the information that the archive preserves. Consumers use the preserved information. A special class of consumers is the Designated Community--the subset of consumers who are expected to understand the archived information. Management is the entity responsible for establishing the broad policy objectives of the archive (e.g., determining what types of information are to be archived, identifying funding sources, etc.). The management entity does not include the day-to-day administration of the archive; this task is performed by a functional entity within the archive itself.

Here are two sample environments, drawn from two real-world archives and described in terms of the OAIS concepts depicted in Figure 1:

Sample Environment I:

Archive: Planetary Data System (planetary science data sets)

Management: National Aeronautics and Space Administration (NASA)

Producers: NASA flight projects

Designated Community: planetary science community

Sample Environment II:

Archive: Electronic and Special Media Records Services Division (U.S. federal records in formats designed for computer processing)

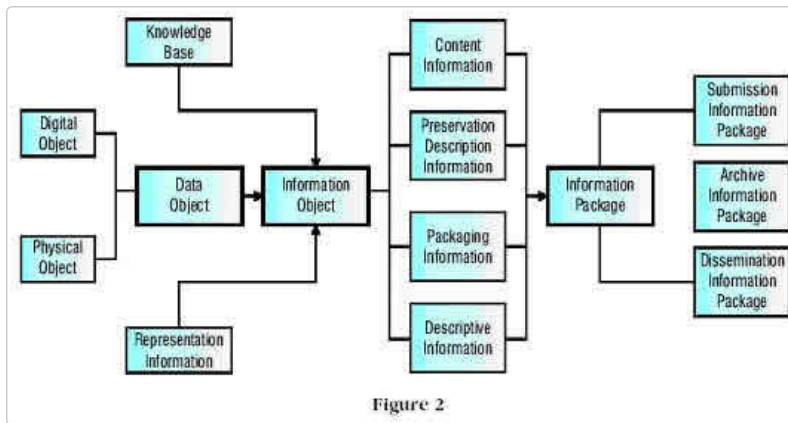
Management: National Archives and Records Administration

Producers: U.S. government agencies

Designated Community: general public

The OAIS Information Model

An OAIS-type archive incorporates the information model shown in Figure 2.



Information is understood to mean any form of knowledge that can be exchanged. In the context of the OAIS, information can exist in two forms: either as a physical object (e.g., a paper document, a soil sample), or as a digital object (e.g., a PDF file, a TIFF file). These two types--physical or digital--may be referred to collectively as the data object.

Interpretation of the data object as meaningful information by the archive's Designated Community is achieved through the combination of the Designated Community's knowledge base, and the representation information associated with the data object. Each individual (or class of individuals, in the case of a Designated Community) has a knowledge base, which is used to understand and interpret information. For example, a Designated Community consisting of Java programmers is expected to have the knowledge base to understand information in the form of Java source code.

The knowledge base of the Designated Community is not always sufficient to fully understand the archived information. In this event, the data object must be supplemented by representation information so the data object can be fully understood by the Designated Community. For example, if the Designated Community consists of all programmers, rather than Java programmers specifically, then information pertaining to Java syntax and programming conventions is necessary for this class of consumers to fully understand the archived data object (Java source code).

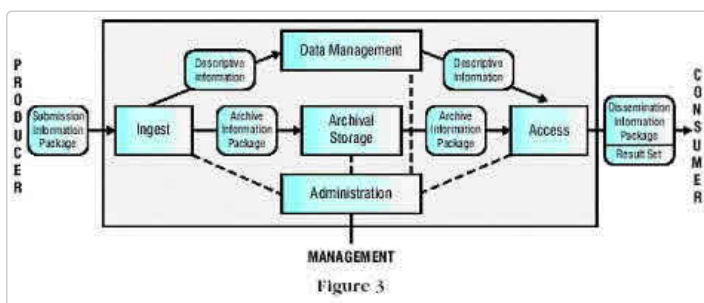
The combination of the data object, the Designated Community's knowledge base, and the representation information results in an information object representing "meaningful information" to the Designated Community. Clearly, meaningfulness is predicated on the definition of the Designated Community the archive serves.

An information package is composed of four types of information objects: Content Information, Preservation Description Information, Packaging Information and Descriptive Information. Content Information is the primary information of interest--the data object and its associated representation information. Preservation Description Information (PDI) contains information necessary to adequately preserve the Content Information it is associated with. In particular, PDI would include provenance information, unique identifiers for the Content Information and information validating the authenticity of the Content Information (such as a checksum or digital signature). Packaging Information binds the components of the information package into an identifiable entity, while Descriptive Information facilitates access to the information package via the archive's search and retrieval tools.

Within the OAIS model, three types of information package are identified: the Submission Information Package (SIP), which is sent from the information producer to the archive; the Archive Information Package (AIP), which is the information package actually stored by the archive; and the Dissemination Information Package (DIP), which is the information package transferred from the archive in response to a request by a consumer.

The Functional Model of the OAIS

Within the OAIS entity (Figure 1), five functional units are identified (shown in Figure 3).



The *Ingest function* is responsible for receiving information from producers and preparing it for storage and management within the archive. More specifically, the Ingest entity accepts information from producers in the form of SIPs, performs quality assurance checks on the SIP, generates an AIP from one or more SIPs and extracts Descriptive Information from the AIPs (metadata for search and retrieval, thumbnail images for browsing, etc.).

Finally, the Ingest function transfers the newly created AIPs to Archival Storage and the associated Descriptive Information to Data Management.

The *Archival Storage function* handles the storage, maintenance and retrieval of the AIPs held by the archive. These responsibilities include receiving new AIPs from the Ingest function and assigning them to permanent storage according to various criteria (media requirements, expected utilization rates, etc.), migrating AIPs to new media as required, error checking, implementing disaster recovery strategies, and providing copies of requested AIPs to the Access function.

The *Data Management function* coordinates the Descriptive Information pertaining to the archive's AIPs, in addition to system information used in support of the archive's operation. In particular, the Data Management function maintains and administers the database containing this information; executes query requests received from the Access function and generates result sets to be returned to the requestor; creates reports in support of the Ingest, Access or Administration functions; and performs updates on the Data Management database, including the addition of new Descriptive Information received from Ingest or new system data received from Administration.

The *Administration function* manages the day-to-day operation of the archive. This includes negotiating submission agreements with information producers and performing system engineering, access control and customer services. The Administration function also performs regular audits of SIPS to assess their compliance with the submission agreement, and develops policies and standards related to the system's data standards (e.g., data format standards, documentation requirements, storage, migration and security policies). This function also serves as an interface between the archive and two components of the OAIS environment: management and the Designated Community (Figure 1).

The *Access function* helps consumers to identify and obtain descriptions of relevant information in the archive, and delivers information from the archive to consumers. This function involves the provision of a single user interface to the archive's holdings for both search and retrieval purposes; generating a DIP in response to a user request by obtaining copies of the appropriate AIP(s) from Archival Storage; obtaining relevant Descriptive Information from Data Management in response to a query; and finally, delivering the DIP or query result set to consumers.

The five OAIS functional entities manage the flow of information from information producers to the archive, and from the archive to consumers. Taken together, they identify the key processes endemic to most systems dedicated to preserving digital information. It is likely that a digital archive will contain functional components similar to those described above, although the specific implementation will differ from archive to archive.

Standardization and AWIICS

The environment, information model and functional entities of an OAIS-type archive interact to form a broad conceptual framework characterizing the primary entities, relationships and processes of an archive dedicated to the preservation of digital information. Implementation of this framework requires the elucidation and integration of standards, policies and procedures that permit the archive to meet its specific objectives, in addition to the OAIS minimum responsibilities listed above.

The OAIS initiative is moving forward with efforts to explore possible areas for standardization within the framework of the reference model. In support of this objective, the Archival Workshop on Ingest, Identification, and Certification Standards (AWIICS) was convened in College Park, Maryland, in October 1999. The purpose of the workshop was to develop an initial agenda for pursuing standardization in the areas of ingest (interaction between the archive and the data producer), identification (establishing a system of permanent, unique identifiers for archived digital objects) and certification (development of accreditation policies, protocols, etc., to establish the authenticity, quality and usefulness of an archive's holdings).

An important goal of the workshop was to determine who among the attendees would be interested in participating in working groups dedicated to pursuing standardization in the areas of ingest, identification and certification. Response from the AWIICS attendees was positive, and groups are expected to form and to begin meeting regularly in the near future.

Why OAIS?

Digital information affects institutions of all kinds, from libraries and archives to corporations and government agencies. A point raised in the reference model documentation is worth repeating here: that digital preservation issues affect all institutions managing information in digital form, including those that do not perceive themselves as performing any type of formal archiving function. The AWIICS conference included representatives from government agencies, libraries, archives, corporations and universities.

Because digital preservation affects such a diverse community, it is useful to distill the issue down to an elemental set of concepts, relationships and processes common to a wide cross-section of digital preservation activities. These reference points serve as the common ground from which joint discussion and mutually beneficial collaboration can proceed. The OAIS reference model elucidates the functions and processes common to nearly all digital preservation environments.

The development of standards in support of the OAIS reference model may serve to promote interoperability among digital libraries, archives and other institutions maintaining digital information over the long term. This is especially significant if it can achieve cooperative efforts between institutions that in the past saw no opportunities for such activity; or, if they did, had no practical means of exploiting them.

Widespread adoption of the OAIS framework could also have potential economic benefits. Standardization across common entities and processes opens the door for cost reduction through shared system components. In addition, standardization promotes the development of broad markets for vendors to support, as systems move from costly customized products and services toward less-expensive standardized versions.

It is not yet clear whether the OAIS initiative will be the consensus approach to the long-term maintenance of digital information. It seems likely, however, that the approach that does emerge, whether based on the OAIS or some other model, will follow a similar development path as the OAIS, with an emphasis on broad participation in open forums. At the very least, the OAIS is laying important foundations for a coordinated and widely applicable solution to the challenges of digital preservation. Active participation in this effort by libraries and librarians will likely yield substantial benefits, both to the library community and to the OAIS initiative.

More Information

The OAIS reference model, proceedings from AWIICS, and other material related to the preservation of digital information is available at the "U.S. Efforts Towards ISO Archiving Standards" Web site at < <http://ssdoo.gsfc.nasa.gov/nost/isoas/us/overview.html>>.

-- **Brian Lavoie** is associate research scientist, OCLC Office of Research.

This article originally appeared in the OCLC Newsletter, No. 243:26-30 (January/February 2000), and is available in its original format at < <http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000001747>> (PDF: 1852K/56pp.)



Follow OCLC
Research:



[_\(https://twitter.com/oclc/lists/oclc-research\)](https://twitter.com/oclc/lists/oclc-research)



[_\(https://www.facebook.com/OCLCResearch/\)](https://www.facebook.com/OCLCResearch/)



[_\(https://www.youtube.com/user/oclcresearch\)](https://www.youtube.com/user/oclcresearch)



[_\(https://www.oclc.org/research/publications/blogs.html\)](https://www.oclc.org/research/publications/blogs.html)

© 2020 OCLC (<https://policies.oclc.org/en/copyright.html>)

Domestic and international trademarks and/or service marks of OCLC, Inc. and its affiliates
(<https://policies.oclc.org/en/copyright/trademarks.html>)

This site uses cookies. By continuing to browse the site, you are agreeing to our use of cookies. See OCLC's cookie notice to learn more. (<https://policies.oclc.org/en/privacy/cookie-statement.html>)