

Plans for an OAIS Interoperable Data Architecture and Protocols

Michael W. Kearney III

Sponsored by Google, Huntsville, Alabama, USA.
KearneySolutions@gmail.com

David Giaretta

PTAB Ltd, Dorset, UK.
David@giaretta.org

John Garrett

Garrett Software, Columbia, Maryland USA
Garrett@his.com

Steve Hughes

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
John.S.Hughes@jpl.nasa.gov



Copyright © 2018 by Michael Kearney, David Giaretta, John Garrett and Steve Hughes. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract

Long-term digital preservation archives are critical to preserving any culture's history and accomplishments. However, these archives typically have unique bespoke access methods for each archive. This paper announces a new initiative to establish an Interoperability Framework for Open Archival Information System (OAIS) archives. We believe a well-understood interface and access method will not only improve efficiency and capabilities for archivists and researchers but also better insure access to such archives in the long term, better enabling the preservation mission of OAIS archives.

This paper provides background on the CCSDS Data Archive Interoperability (DAI) Working Group that established the OAIS reference model and other associated standards. It describes the plan to extend the OAIS ecosystem to include an OAIS Interoperability Framework (OAIS-IF) and the software architecture that enables it. The architecture includes producer and consumer interfaces, an

abstraction layer and “plug-ins” or “bindings” (APIs or protocols) that handle the unique needs of access to each discipline archive.

We are presenting this paper with the goals of (1) increasing awareness of OAIS and its applicability for digital preservation archives, (2) explaining the plan for an interoperable data architecture applicable to most preservation archives, (3) seeking peer review for this approach, and (4) explaining how archive providers might play a role in developing a part of this architecture specific to their archive instantiation or discipline. Specifically, we anticipate that the standards and prototypes for some of these bindings will be provided by the discipline communities or by developers of OAIS archives to serve their specific needs.

Keywords: CCSDS, OAIS, Archive, Architecture, Standards

Introduction

The Consultative Committee for Space Data Systems (CCSDS) began in 1982¹ as a body of space agencies coordinating agreements on data and communications systems in order to foster interoperability between agencies’ systems, hence enabling jointly cooperative spaceflight missions among the participating agencies. While organized by the governments of the participating space agencies, it is inclusive of commercial, academic and other organizations that develop or participate in space programs. Within CCSDS, the Data Archive Interoperability (DAI) Working Group (WG) team has been working on digital preservation standards since their earliest discussions in 1995 (under predecessor group names), and initiated the first publication of the Reference Model for OAIS in 2002.

CCSDS also functions as a subcommittee under the International Organization for Standardization (ISO); Technical Committee 20, Subcommittee 13 (TC20/SC13), *Space Data and Information Transfer Systems*. In this capacity, CCSDS recommended standards are processed as ISO standards, and review is solicited from the ISO community or discipline that would be interested. Of all the CCSDS working groups, the DAI WG has the broadest participation from the ISO community and non-space organizations.

The DAI WG first addressed the broadest problems of processes and practices for systems and organizations that are necessary to establish trusted digital archives. The result was the [Reference Model for an Open Archival Information System² \(OAIS\)](#). While their charter was to address the digital preservation problem for the CCSDS space agencies, the DAI WG attacked the “root cause” of the fundamental problems in a very generic way. The combination of that approach and the very broad participation of the ISO community, national archives and libraries caused the OAIS Reference Model standard to “go viral” and become the most broadly accepted digital preservation archive standard, practiced in archives and libraries around the world.

Additionally, [CCSDS³](#) and [ISO⁴](#) standards for *certification* of trustworthy repositories have been developed, and resulting processes and [organizations](#) for certification have matured. They are based on OAIS and they have established a sustainable and healthy environment for long-term digital preservation.

However the primary focus for the DAI WG until now has been *processes and practices*. Implementable standards for OAIS archives was intentionally not addressed so that the process framework would have the broadest applicability regardless of implementations, hence

fostering a widely accepted methodology to establish and operate trusted repositories for preserving the world's digital heritage.

A few years ago, Vint Cerf, Google's Vice President and Chief Internet Evangelist, approached participants in the DAI WG and advocated that while the OAIS provided a great Process Framework for digital preservation, the lack of technical (implementable) standards addressing access and interoperability for long-term preservation archives was a key problem for sustained access to those archives. After some discussion within the WG, a concept for future work in this area was proposed to the CCSDS Management Council, and the proposal for work in this area was approved.

To complete a healthy OAIS ecosystem, interoperable access for digital archives should be established with implementable specifications. These standards would allow common user interfaces to be interoperable with various archives that allow external users to access the valued resources and information within OAIS archives. If this becomes broadly accepted by the user communities and well-understood through future generations, it will contribute to longer-term accessibility of those preserved digital assets. Besides user interfaces, interoperable archive-to-archive access may also be addressed, if it is also encoded in the OAIS Process Framework.

The term being used to frame this new effort for technical and interoperable standards is the *OAIS Interoperability Framework* (OAIS-IF). If this interoperable access scheme is successful and is as broadly accepted as the OAIS Reference Model, then the possibility of ubiquitous access to long-term preservation archives will reach not only to archives around the world but to distant generations of future researchers. Therein lies the goal of this new effort in the CCSDS DAI WG.

The goal of this paper and associated presentation is to solicit both peer review of the concepts herein and participation in the DAI WG effort by contributors that can address the technical issues of OAIS-IF interoperable Information Technology (IT) systems.

OAIS Background

An OAIS is an archive, consisting of an organization of people and systems that have accepted the responsibility to preserve information and make it available for a Designated Community. The Reference Model for OAIS is published as both CCSDS and ISO standards (as ISO adopts most CCSDS technical standards). OAIS has global recognition and adoption as described in [Wikipedia](#)⁵: “The OAIS has become the standard model for digital preservation systems at many institutions and organizations. OAIS compliance has been a stated fundamental requirement for major digital preservation efforts at the US National Archives and Records Administration⁶, the US Library of Congress, The British Library, Bibliothèque Nationale de France, The National Library of the Netherlands...” (And many others.) The CCSDS Space Agencies have some examples of successful implementations of OAIS processes in NASA's Planetary Data System (PDS)⁷ and the National Space Science Data Center (NSSDC) both of which have international participation and contribution.

OAIS Relationship with the proposed new architecture

An OAIS archive aims to be trustworthy by implementing the OAIS Reference Model and other documents of the OAIS Process Framework (OAIS-PF). An OAIS archive that

implements the OAIS Interoperability Framework (OAIS-IF) also aims to be *interoperable* with other entities that implement or use the OAIS-IF standards.

The OAIS Process Framework (OAIS-PF), including both informative and normative documents, will remain untouched by this new OAIS-IF effort. The OAIS certification process (as documented in CCSDS 652.1 and ISO 16363) is well established and need not be modified to accommodate the OAIS-IF standardization initiative. The OAIS Reference Model and associated documents of the OAIS-PF provide a coherent system of practice which structure the behavior needed for a digital repository to be trustworthy. The OAIS RM does not, however, provide interoperability between users and archives or between archives. OAIS is intentionally broad and accepting of all implementations, because it enables confidence and trustworthiness in any archive that follows OAIS practices, regardless of implementation.

The OAIS Interoperability Framework (OAIS-IF) is an additional tool set available to archives to enhance an OAIS-compliant archive by adding capabilities for interoperability between users and archives, or between archives. The OAIS-IF will support OAIS compliance because it will be a suite of standards that provide interface and protocol functions for conducting the interactions required by an OAIS archive. To enforce interoperability, the OAIS-IF will necessarily impose some constraints (but not total constraint) on implementations of archives. Primarily, these constraints will be at the level of communications, interaction patterns, protocols and interfaces.

So we have two suites of standards in the OAIS ecosystem. The OAIS-PF is the basic suite of standards (Ref Model, Certification, etc.) that guides all trustworthy repositories, whether interoperable or not. OAIS-IF is an additional suite of standards that enable interoperability for those archives that choose to add it to their capabilities, or for those archives whose users have demanded an interface that is more compatible and interoperable with their systems and with other interoperable archives.

The Proposed OAIS Interoperability Framework

The first OAIS-IF product that will be produced by the DAI WG is the Data Archive Architecture Description Document (DA-ADD). The ADD will describe the overall framework from top to bottom. It is a prerequisite to insure a compatible end-to-end design of the other components in the OAIS-IF architecture.

The framework consists of an upper and middle framework that remains relatively consistent, with a lower framework of unique plug-in drivers (APIs or protocols) that adapt the framework to each unique archive. This is illustrated in Fig. 1.

The upper framework provides the user interfaces for the producers and consumers of the OAIS archive data. The intention is to engineer these user interfaces so that they channel data to/from many types of archives with little or no adaptation required for various archive types.

The middle framework is an Archive Abstraction Layer (AAL) whose function is to hide the implementation details (the mechanics of a specific archive's characteristics) to facilitate interoperability and archive platform independence. Example interactions that pass through the AAL would include methods to (for example):

- Allow a data producer to send data to the archive in the form of a Submission Information Package (SIP);

- (If salient at the archive interface) perform any functions needed to support the archive’s creation of Archival Information Packages (AIPs), the key OAIS component for long term preservation;
- Allow a consumer to query an archive about the data objects available in the archive and about metadata associated with the preserved data;
- Allow a user, via the user interface, to select data and metadata to be retrieved from the archive.
- Allow a data consumer to request and receive a Dissemination Information Package (DIP)

It is not certain at this time whether the characteristics of the archive’s internal AIP will “surface” at its interface to the OAIS-IF. The AIP encapsulates all the essential pieces of information needed to enable preservation of the data. This would include Representation Information and other metadata. However, while OAIS requires AIP functionality, the entire interoperability framework functionality may be simplified to handling SIPs and DIPs, and it may never be exposed directly to AIPs. This will become clearer as the architecture matures.

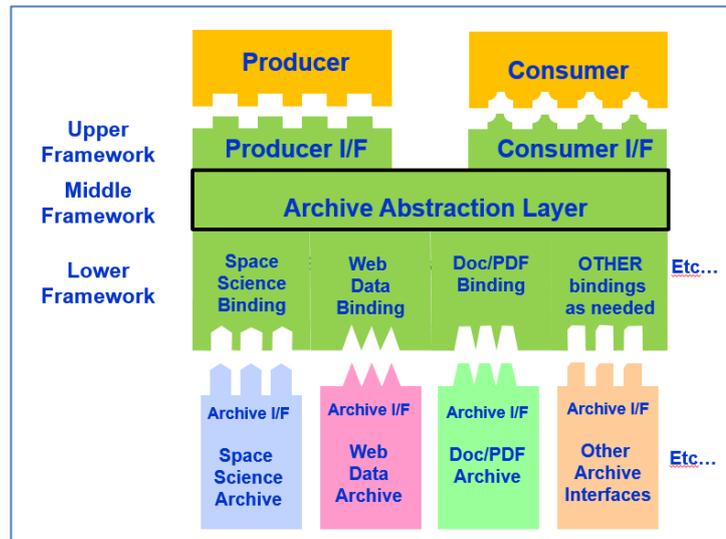


Fig. 1: The overall high-level structure of OAIS-IF

Finally, in the lower framework are the specific plug-in components that connect the Interoperability Framework to each archive. The exact nature of these components is not yet defined; they may be protocols or APIs or bindings in the same manner that the CCSDS SM&C WG’s Message Abstraction Layer (MAL) has bindings to various communications functions. For now we will refer to these generically as bindings, the term more generally used in this environment. The preferred fidelity of the breakdown of these components may vary from implementation to implementation. For example, spaceflight missions may have bindings for each of various mission functions (telemetry, spacecraft housekeeping, engineering or enterprise data), or for each of various scientific disciplines (astronomy, earth observing, materials research, genomic research, etc.). However, there could be greater efficiency in an all-encompassing binding for all forms of telemetry. For other disciplines there could be bindings based on media format (audio, video, documents). There could be dedicated bindings for specific archives or organizations with many archives (NASA, National Archives, Libraries, Cloud services, etc.) or product vendors for their file types (PDF, Oracle, Google Docs, etc.). While the CCSDS DAI WG will standardize the framework and drivers for the space agencies, we expect that those other outside organizations or product vendors will develop the standards (and bindings) for their own organization, product, or discipline.

If your organization anticipates that you may want to develop an OAIS-IF interoperability binding standard or implementation for your discipline to use with your archive(s) or to support cross-discipline research, it might be a good idea to consider joining the CCSDS DAI WG now, to better understand the architecture that will service your archive, as well as to generally become part of the future of digital preservation.

OAIS-IF Capabilities

This OAIS-IF scheme provides several kinds of interoperability. At the top layer, broad acceptance of this framework would mean that many researchers could utilize one framework and common tools for access to many of their discipline's resources. At the bottom layer, as long as organizations make the drivers widely available (downloads of plug-ins from their websites) then archive researchers will have access to many archived resources outside their traditional domain. This is because a researcher in a specific binding (for example archaeology) should be able to "plug in" a binding from another discipline (for example, genome research) into his installation with his familiar consumer interface and abstraction layer. At that point the researcher should be able to access other disciplines' archives depending on what bindings he has access to. Cross-discipline research is expected to be a major driver of worldwide technology advancement in the future "big data" era, and this innovative data architecture for archives may well be a leading enabler of that cross-discipline research capability for the long-range future.

And of course with the OAIS requirements for trustworthy repositories factored into the functionality of this software architecture, the OAIS-IF supports the functions that are essential to long-term digital preservation.

The architecture enables compliant interoperable archives to perform the core functions of OAIS archives. The basic functions of an OAIS are to import a Submission Information Package (SIP), compose a resident Archive Information Package (AIP), and export a Dissemination Information Package (DIP). These functions are defined and described in detail in the OAIS Reference Model standard. They are illustrated in Fig. 2.

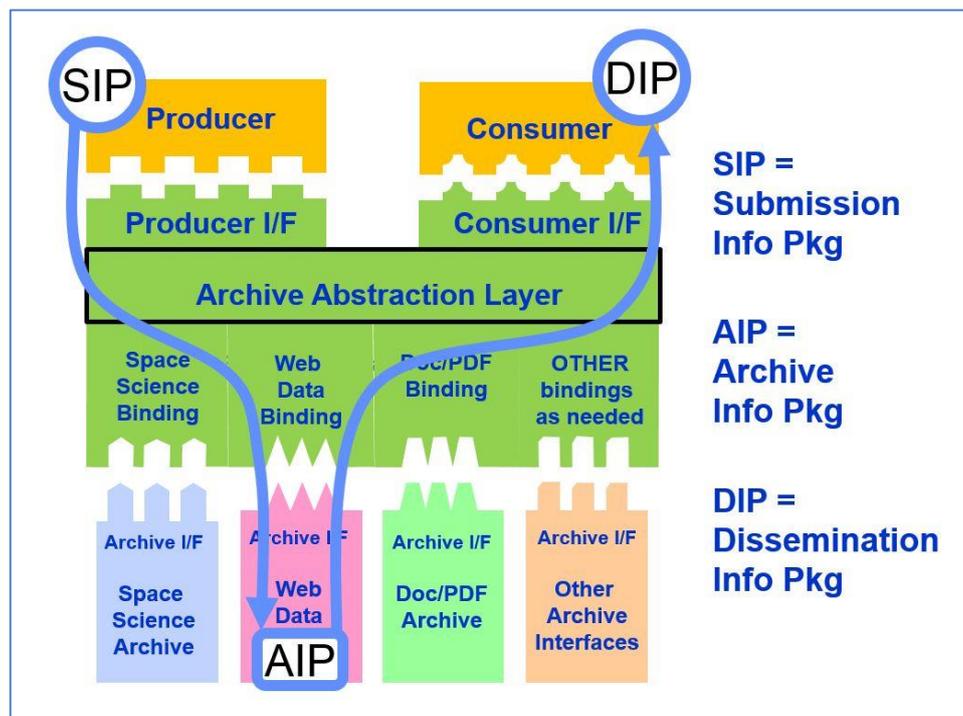


Fig. 2: OAIS Archive Performing SIP, AIP and DIP functions.

Standards to be produced

The specific standards that are part of the current forward DAI WG work plan are:

- DA-ADD – Data Archive - Architecture Description Document
- PAIS – [Producer Archive Interface Specification](#) (already produced⁸ but may be revised)
- PAIP – Producer Archive Interface Protocol
- CAIS – Consumer Archive Interface Specification
- CAIP – Consumer Archive Interface Protocol
- AAL – Archive Abstraction Layer
- Bindings – specific to each archive or domain or format.

For the bindings, the CCSDS DAI WG plans to produce standards and prototypes for bindings (plug-ins/protocols/APIs) that support the missions of the CCSDS participating Space Agencies. Other bindings for outside the space industry will be produced by organizations that the DAI WG recruits to produce them, associated with a specific archive, discipline, product format or other domain that is specific to the expertise in that organization. And of course, once the OAIS-IF standards garner broad acceptance, we expect that organizations will initiate development of these tools for their community or customers' access on their own without needing solicitation or coordination.

Fig. 3 illustrates where these standards fit in the OAIS-IF architecture. Note that the Architecture Description Document (ADD) encompasses the entire architecture and establishes the relationships of the other standards.

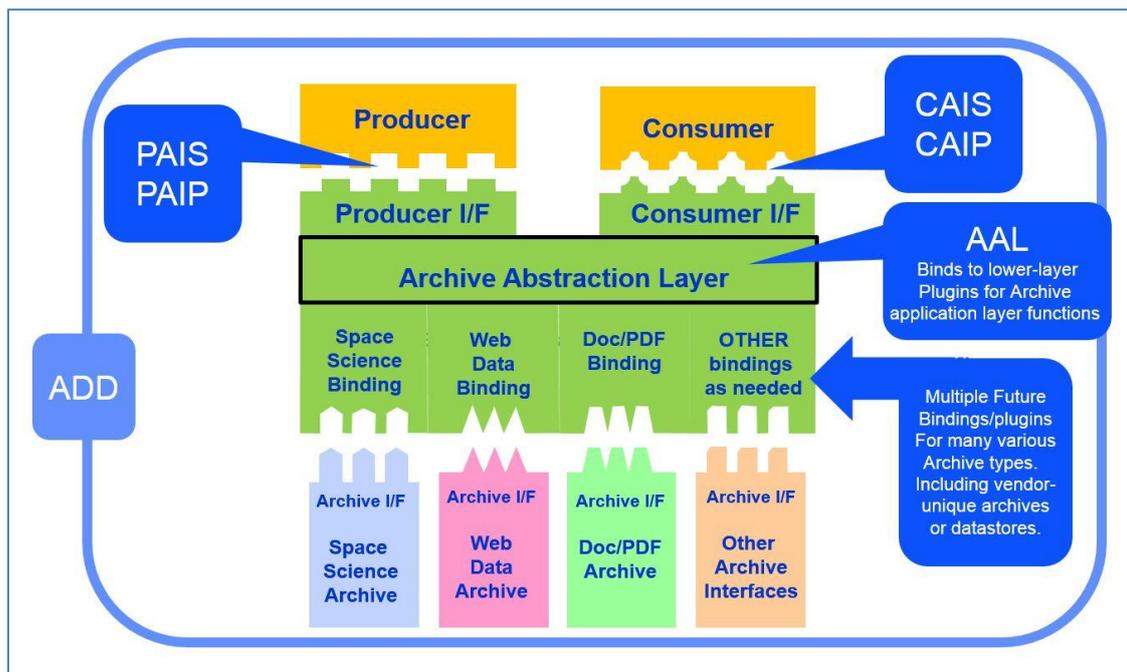


Fig. 3: Standards as Elements of the OAIS-IF and Where They Fit in the Architecture

Integration with OAIS structures and functions

Fig. 4 provides a more detailed view with specific OAIS functions and relationships to the communications layers. This illustrates that the bindings for the various archives, data types, or functions will provide OAIS capabilities for provenance, reference information, fixity and

other OAIS and AIP functions. These functions will be accessible by users or user applications at the top layer of the OAIS-IF architecture.

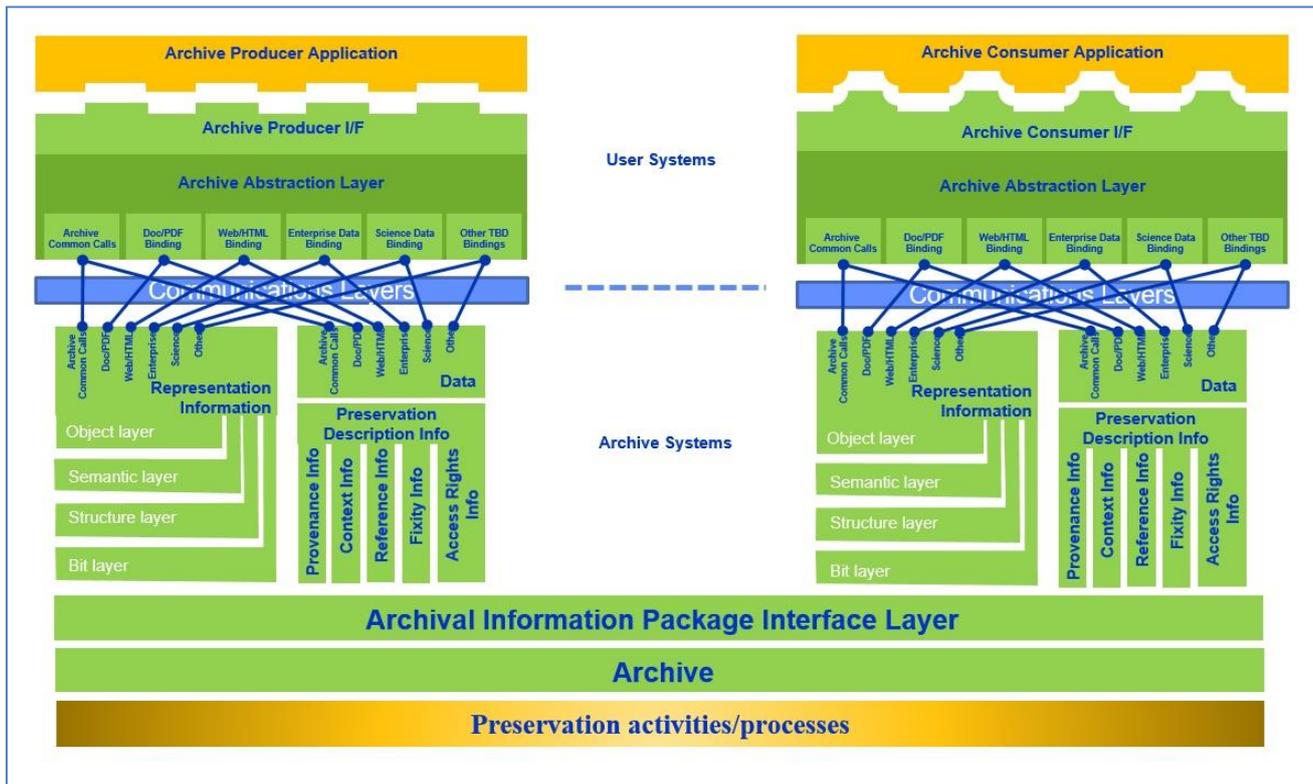


Fig. 4: OAIS-IF Integration with OAIS Functions and Communications Layers

Open Work for OAIS-IF

There are several aspects to a software architecture of this nature that have yet to be addressed for the OAIS-IF. Of course, security factors will have to be designed in to this architecture from the outset. There are some unique problems of archives that will need to be addressed, such as protection of intellectual property, etc. Many of these considerations are already addressed in OAIS as practices and processes, but the OAIS-IF architecture and protocol standards will need to encode those OAIS processes in system interface design. The overall concept and the use of the abstraction layer is similar to initiatives in other CCSDS areas (Planning and Mission Ops Services also use a Service Oriented Architecture with an abstraction layer), so the design has a developing heritage of broader usage.

Challenges ahead for OAIS-IF

The OAIS-IF is based on a concept that many different kinds of archives and designated communities can be supported with a common user interface that “channels” discipline-specific responses from the archive to the user or their application. The goal is to “quarantine” the archive-specific and discipline-specific functions in the development of the binding for unique archives/disciplines/formats. The output of those discipline-specific functions are then channeled through the abstraction layer and the user interface for interpretation by the user at that user interface. Some peer reviewers have expressed skepticism that this may not be achievable for all of the archives served by OAIS now and in the future. The concern is that the Designated Communities require unique capabilities at their archive interfaces that cannot be accomplished with a generalized interface capability. The team that is working on the

OAIS-IF standards believes that with the right approach to presenting archive functions to the user, almost all scenarios can be accommodated. However, a fallback plan would be to provide capabilities for the generic standardized user interface to coexist with a small discipline specific function for the rare cases where there are some remaining discipline-unique functions that cannot be accommodated by the generic user interface. That fallback plan is not being enacted at this time. Also, in an extreme case, legacy interfaces can coexist with the new OAIS-IF capabilities for as long as needed, until the legacy users are no longer driving the need for bespoke capabilities.

Compatibility with existing archive infrastructure

When existing OAIS archives evolve to support the new interoperability functions that are provided by OAIS-IF, the existing bespoke legacy interfaces for data producers and data consumers can function in parallel to the same archive as long as necessary. Fig. 5 illustrates a scenario where OAIS-IF is added to an existing archive while the current or legacy archive interfaces are retained. This can be a long-term situation if current users are reluctant to give up a familiar interface while the archive is providing OAIS-IF interfaces to benefit new users

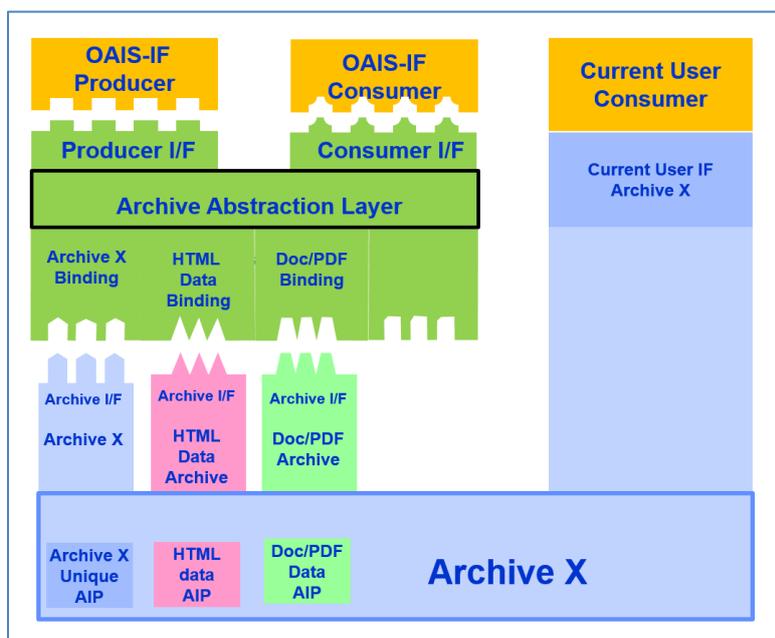


Fig. 5: OAIS-IF compatible with current/legacy interfaces

The overall OAIS-PF and OAIS-IF environment

Now to back out and look at how this fits with the overall OAIS environment. Fig. 6 frames the environment between data producers, data consumers and data preservers (the archives). In this illustration, the crisp division between the OAIS-PF and the OAIS-IF illustrates that while the OAIS-IF supports all OAIS functions, the OAIS-IF is not required for the archive to achieve the status of an OAIS trusted repository. An OAIS repository can perform all required functions of a compliant trustworthy repository without the OAIS-IF. However such a repository will be missing out on the broader user access and interoperability capabilities that come with OAIS-IF standardization.

and cross-domain researchers, etc. Or it can be simply for a short transition period so that the current access and the new access can be “field tested” in parallel for as long as needed.

This approach is similar to the legacy interface approach that can coexist with the new OAIS-IF capabilities discussed in the previous section. The OAIS-IF adds functionality for new interoperable capabilities and enhanced access, but does not require existing functionality to be deprecated.

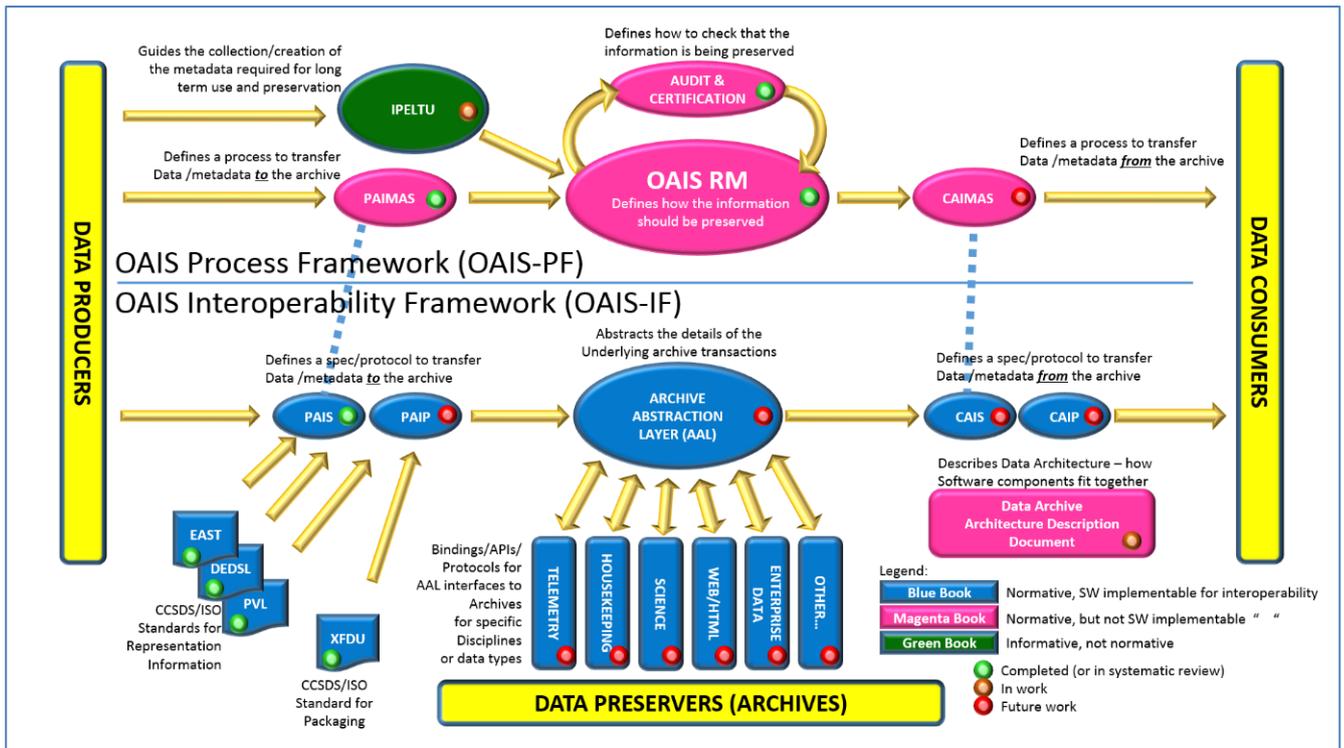


Fig. 6: Overall OAIS, illustrating the OAIS-PF and OAIS-IF Environments

Conclusions

- (1) OAIS processes assure trustworthiness of digital archives. The addition of an OAIS Interoperability Framework will bring benefits of greater efficiency, cross-domain capability, and more assured long-term preservation to digital preservation archives around the world.
- (2) While there are other efforts to produce “standard formats” for some archives, there is no other work that we know of that uses the innovation of unique plug-in protocols for each designated community, or data format, or organization. And it is the only effort using the innovation of an abstraction layer and an architecture that enables cross-discipline research.
- (3) Preserving the history in our digital assets is critically important, because all history is and will be digitally encoded for the current and foreseeable generations; very little critical history is entrusted to paper or clay tablets any more. It is essential that we solve the “digital dark ages”⁹ problem that results mainly from software evolution and obsolescence.
- (4) This is relevant to all archives endeavoring to provide long-term preservation of their valuable and historic digital assets. If this approach is successful and ubiquitous across our world, one can envision a “big data” future, a century from now, where huge archives of well-preserved historic data are interrogated by AI data miners, and they are missing only a small amount of historical data from the late 20th and early 21st century before OAIS and the OAIS-IF became prevalent.

This architecture is, in a few words, all-inclusive, all-serving, OAIS-compatible and modular. It enables cross-discipline research and cross-organizational interoperability. The DAI Working Group welcomes review and critique, and participation in the development of OAIS-IF from users or developers that can make substantive technical contributions.

Repeating this statement for emphasis: If your organization anticipates that you may want to develop an OAIS-IF interoperability binding standard or implementation for your discipline to use with your archive(s) or to support cross-discipline research, it might be a good idea to consider joining the CCSDS DAI WG now, and become part of the future of digital preservation.

Appendix A – Acronym List

AAL	Archive Abstraction Layer
ADD	Architecture Description Document
AIP	Archival Information Package
API	Application Programming Interface
CAIMAS	Consumer Archive Interface Methodology Abstract Standard
CAIP	Consumer Archive Interface Protocol
CAIS	Consumer Archive Interface Spec
CCSDS	Consultative Committee for Space Data Systems
DAI WG	Data Archive Interoperability Working Group
DEDSL	Data Entity Dictionary Specification Language
DIP	Dissemination Information package
EAST	Enhanced Ada SubseT (Data Description Language)
I/F	Interface
IF	Interoperability Framework
IIPC	International Internet Preservation Consortium
IPELTU	Information Preservation to Enable Long Term Use
ISO	International Standards Organization
IT	Information Technology
JPL	Jet Propulsion Laboratory
MAL	Message Abstraction Layer
NASA	National Aeronautics and Space Administration
NSSDC	National Space Science Data Center
OAIS	Open Archival Information System
OAIS-IF	Open Archival Information System – Interoperability Framework
OAIS-PF	Open Archival Information System – Process Framework
OMG	Object Management Group
PAIMAS	Producer Archive Interface Methodology Abstract Standard
PAIP	Producer Archive Interface Protocol
PAIS	Producer Archive Interface Specification
PDF	Portable Document Format
PDS	Planetary Data System
PVL	Parameter Value Language
RM	Reference Model
SC13	Subcommittee 13
SIP	Submission Information Package
SM&C WG	Spacecraft Monitor and Control Working Group
TC20	Technical Committee 20
WG	Working Group
XFDU	XML Formatted Data Unit

Acknowledgments

The authors wish to acknowledge the CCSDS organization for the space agencies' vision in establishing a premier Voluntary Consensus Standards (VCS) development organization which supports and encourages critical foundational work such as that done in the DAI Working Group and other CCSDS working groups.

The primary author wishes to acknowledge Google for sponsoring attendance for digital preservation outreach at forums such as this.

References

- ¹ About CCSDS <https://public.ccsds.org/about/default.aspx>
- ² CCSDS, Reference Model for an Open Archival Information System, <https://public.ccsds.org/Pubs/650x0m2.pdf>
- ³ CCSDS, Audit and Certification of Trustworthy Digital Repositories, <https://public.ccsds.org/Pubs/652x0m1.pdf>
- ⁴ ISO. ISO-16363:2012, Audit and Certification of Trustworthy Digital Repositories, <https://www.iso.org/standard/56510.html>
- ⁵ Wikipedia, Open Archival Information System, https://en.wikipedia.org/wiki/Open_Archival_Information_System
- ⁶ US National Archives and Records Administration (NARA), Strategy for Preserving Digital Archival Materials, <https://www.archives.gov/preservation/electronic-records.html>
- ⁷ Hughes , J.S., Crichton, D., Hardman, S., Law, E., Joyner, R., Ramirez, P., PDS4: A Model-Driven Planetary Science Data Architecture for Long-Term Preservation, IEEE 30th International Conference on Data Engineering (ICDE), Chicago, IL USA, 2014.
- ⁸ CCSDS, Producer-Archive Interface Specification (PAIS), <https://public.ccsds.org/Pubs/651x1b1.pdf>
- ⁹ Wikipedia, Digital Dark Age, https://en.wikipedia.org/wiki/Digital_dark_age