A New Future for Archive Interoperability and Preservation

Mike Kearney[a]*, David Giaretta[b], John Garrett [c], and Steven Hughes[d]

[a] *Space Infrastructure Foundation, Consultant to Google, Huntsville, AL 35803, USA*

[b] *PTAB Ltd, Dorset, DT9 5LP, UK*

[c] *Garrett Software, Columbia, MD, 21044 USA*

[d] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109 USA*

* Corresponding Author

Abstract

The purpose of this paper is to provide updated information on the OAIS-IF project. The purpose of the OAIS-IF project, developed by the CCSDS (Consultative Committee for Space Data Systems) Data Archive Interoperability (DAI) working group, is to provide interoperability between archive users (producers and consumers of archive data) and the archives, as well as archive-to-archive interoperability.

CCSDS is the leader in data and communications standards for space missions. The DAI working group is their team for long-term digital preservation standards. This paper will first provide background on the CCSDS, the DAI team, and the Open Archival Information System (OAIS) Reference Model (RM). It will then introduce the concept of an OAIS Interoperability Framework (OAIS-IF); a suite of standards that will support the OAIS and add capabilities for interoperability between users and all archives that comply with the OAIS-IF standards.

Interoperability was at the core of the OAIS-IF requirements from the outset. However, (since the presentation at SpaceOps 2018) the team has revamped the data architecture of the OAIS-IF, providing greater modularity, greater capabilities for cross-discipline research, and support for non-OAIS archives. This is the focus of this manuscript. This will benefit space mission archives, long-term preservation archives, and any other archives that choose to implement OAIS-IF.

The technical concept is based on adapters at both the user location and the archive location that will provide access to all OAIS-IF compliant systems. A user will require only one adapter set to access archives of many disciplines. And an archive will require only one adapter set to serve users of many disciplines. The "adapter set" on each side consists of a generic adapter and a specific adapter. The generic adapter provides basic functions which can serve many users and many archives. The specific adapter on the user or archive side performs functions specific to the archive, if needed. The result is software reusability for the generic adapters, reducing development burden.

This will provide a capability for interoperability between the many users and archives that will eventually subscribe to these standards. If the OAIS-IF achieves ubiquitous acceptance by digital preservation archives, as was achieved by the original OAIS standard, the result will be improved cross-discipline and long-term accessibility and usability, clearly benefiting the goals of long-term digital preservation and the goals of our spaceflight missions, thereby increasing the value of the missions through which the information which has been created and captured.

Keywords: Archive, Preservation, OAIS, Interoperability, Framework

Acronyms/Abbreviations

See Appendix A.

1. Introduction

The primary motivation for these co-authors to publish this work is to introduce the space community to this breakthrough work in archive interoperability. We hope to kindle an interest in the community so that we will receive peer review input on our techniques and technology for developing the standards and to encourage adoption after the standards are completed.

17th International Conference on Space Operations, Dubai, United Arab Emirates, 6 - 10 March 2023.

Additionally, we hope to promote the importance of long-term digital preservation among the space community. Especially in this era when large historic datasets are becoming valuable for Artificial Intelligence (AI) analysis, the value of long-term digital preservation of most program data (not only science data) cannot be overstated.

2. CCSDS and the DAI Working Group

CCSDS began in 1982 [1] as a body of space agencies coordinating agreements on data and communications systems in order to foster interoperability between space and ground 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, and others who use the results from those programs.

The Data Archive Interoperability (DAI) Working Group (WG) [2] has been working on long-term digital preservation standards since their earliest discussions in 1995, and the first publication of OAIS in 2002. The DAI WG is within the Mission Operations and Information Management Services (MOIMS) Area of CCSDS, which is basically the applications layer of the CCSDS architecture. **Error! Reference source not found.** illustrates the high-level end-to-end architecture of CCSDS and the yellow arrow highlights where the DAI WG fits in.

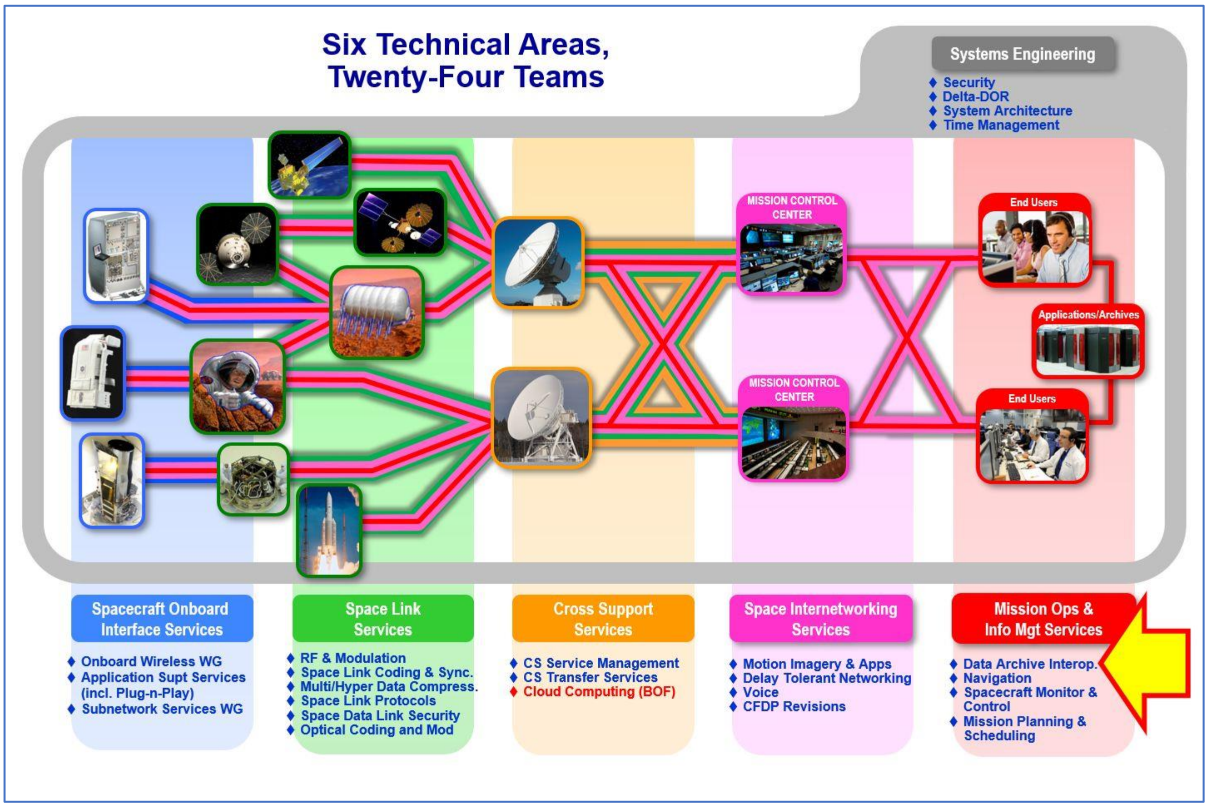

Figure 1: The CCSDS End-to-end Architecture illustrating where the DAI WG Fits in.

3. Open Archival Information Systems (OAIS) background

The DAI team first addressed the broader problem 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 System (OAIS) [3]. Because CCSDS also functions as an ISO subcommittee (ISO TC20/SC13) [4], the OAIS Reference Model was also issued as an ISO standard (ISO 14721) [5]. Eventually ISO-compliant certification standards and processes were set up (ISO 16363 [6] and ISO 16919 [7]), and a certification system [8] is now operational. Did you know that your OAIS-compliant archive can be certified as trustworthy?

Eventually the OAIS became the most broadly accepted digital preservation archive standard, adopted as a "de facto" [9] standard practiced in archives and libraries around the world, well outside of the spaceflight community. Many archives, libraries, and repositories are based on OAIS and they have established a sustainable and healthy environment for long-term digital preservation, albeit largely outside the space industry. However, the primary focus until now has been standards for ***processes and practices***. Collectively these are referred to as the ***OAIS Process Framework (OAIS-PF).*** Implementable standards for OAIS archives were intentionally not addressed so that the process framework would have the broadest applicability, hence fostering a widely accepted methodology for trusted repositories.

Note that the references in this paper point to the currently published version of the Reference Model for OAIS. However, a newer version of the OAIS-RM has been approved by CCSDS and is awaiting completion of the ISO review. The newer version is the version that OAIS-IF is synchronized with.

4. A New Direction for the DAI Working Group

Google's Vint Cerf [10] approached us and advocated a stronger focus on technical interoperability for preservation archives. We realized that, to complete a healthy OAIS ecosystem, interoperable access for digital archives is needed, specified in standards, implementable for interoperability. 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 archives. If these standards become broadly accepted by the user communities and the underlying concepts remain well-understood through future generations, it will contribute to longer-term accessibility of those preserved digital assets. These standards will support and draw on the OAIS Reference Model, but it will not require OAIS compliance. The interoperability specifications will be useful as well to non-OAIS archives, and will enhance cross-discipline research.

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.

5. The Digital Preservation Problem Statement for Space

Our previous SpaceOps 2018 paper [11] gave a number of examples of lessons learned from digital preservation failures (Apollo, ALSEP, NASA Galileo, IMAGE) which we won't repeat here. There are even some recent examples of ISS data from early in the program that is not retrievable due to obsolete data formats. However, it is difficult to research such failures after programs have terminated because personnel are dispersed, and even when contacted they often don't want to talk about lost data. In many cases, after mission termination, there is simply not funding allocated for an effort to preserve data with the necessary metadata to make it interpretable by future systems. In some cases, the science return data is preserved in an institutional facility like the Planetary Data System (PDS) that is mostly OAIS compliant, but other program data (engineering test data, program/project financial data) is almost never properly preserved.

A new use case for preservation; The Solar and Space Physics Decadal Survey is currently considering an Interstellar Probe mission [12] that could have a mission duration of at least 50 years. The team operating at the final phase of the mission will need to have access to early mission data. As software obsolescence forces data format changes, it is planned that the historic data would be migrated to new formats. However, if the historic data formats need retrieval, those data will need to comply with the OAIS processes for preserving the information, and their mission will benefit from the OAIS-IF interoperability standards that are currently being developed.

Imagine a large historical archive of not only science data but your agency's or company's past project data. You can ask your AI "For all projects that use cost model X, what percentage of them came in at or under predicted costs?" If you don't have that historic project archive yet, there is no better time to start than now. But if you don't use OAIS practices or interoperability standards, you may lose that capability.

6. OAIS-IF Development Plans and Constraints

OAIS-IF will define interfacing standards only. We do not plan to specify the underlying functionality of software beneath the interfaces, except as needed to enable the interface. We expect to define protocols and Application Programming Interfaces (APIs) which can be used to create software adapters to enable interoperability.

We are developing OAIS-IF using Model Based Engineering (MBE), specifically UML models. In parallel with the published standard (a standalone implementable specification) we also expect to be able to post the UML models online to help developers get a rapid start at developing systems in accordance with the OAIS-IF standards.

CCSDS requires two independently developed interoperable prototypes before a standard can be approved and published. We also hope to make these prototypes available as open source, as references for archive developers.

17th International Conference on Space Operations, Dubai, United Arab Emirates, 6 - 10 March 2023.

7. Interoperability for OAIS-IF

What exactly do we mean by interoperability for archives? We are addressing interoperability for transactions between users and archives, and also for transactions between multiple archives (as in the case of transferring preserved assets from one archive to another.)

OAIS defines two kinds of users [13]. Producers provide data to be stored in an archive, and Consumers withdraw that data for interpretation and usage. Figure 2 illustrates the basic OAIS-IF interface concepts.

We plan for archive-to-archive interfaces to utilize the same interfaces as users. This should enable transfers between archives of preserved data and the supporting metadata that enables its interpretation.

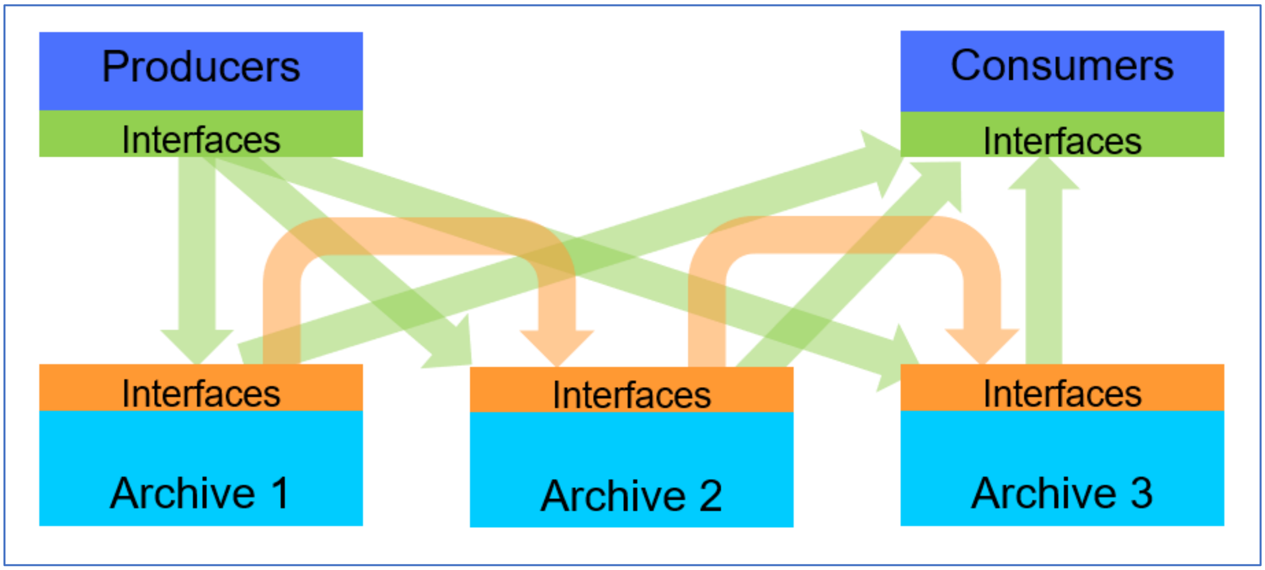

Figure 2: Interoperability Goals for OAIS-IF

8. OAIS-IF Application of Adapters

Our vision is that each client should need only one adapter at their location which consists of a generic part and a part specific to the user's native software. Each user should be able to talk to many archives via their adapter, one installed at each of the archives. And each archive's adapter should be able to talk to many users' adapters. The key point is that instead of requiring the users to have lots of versions of software tailored to each archive to access many different archives, they only need a single adapter. Hence users have access to much more archive data than they had before. This is illustrated in Figure 3.

To describe the Information Objects that are exchanged, we must first explain some terminology. An Information Package consists of the Information Object [14] plus packaging information that allows software to identify (and "unpack") the components of the package. The architecture will support the exchange of Information Packages, as described in the Information Model [15] that is part of the OAIS Reference Model and its updates. For OAIS systems, the OAIS Information Packages consist of a Data Object (which could be the original "target of preservation"), and the Representation Information (RepInfo) that is needed to understand or use it, as well as optional Preservation Description Information (PDI) which consists of Provenance, Reference, Fixity, Context, and Access Rights Information. Provenance in particular is important for proof of authenticity, especially when the information is the basis of policy decisions, and also for reproducibility and replicability, which are needed for a valid scientific process. In OAIS terms, RepInfo is the information needed to understand the Data Object. PDI is the specific type of metadata that are needed to preserve the target of preservation. So, in our architecture, what are exchanged between adapters are Information Packages, which are special types of Information Objects.

While compliance with the OAIS Reference Model will require adequate PDI, archives that are not OAIS compliant can make use of these same OAIS-IF interfaces. In their case, the PDI may be a "null set" or partial set of PDI in an exchanged information object. This allows current non-OAIS archives to achieve the same level of interoperability, depending upon the amount of Representation Information that is available. However, it positions non-OAIS systems to eventually be able to add PDI and grow towards OAIS practices of digital preservation.

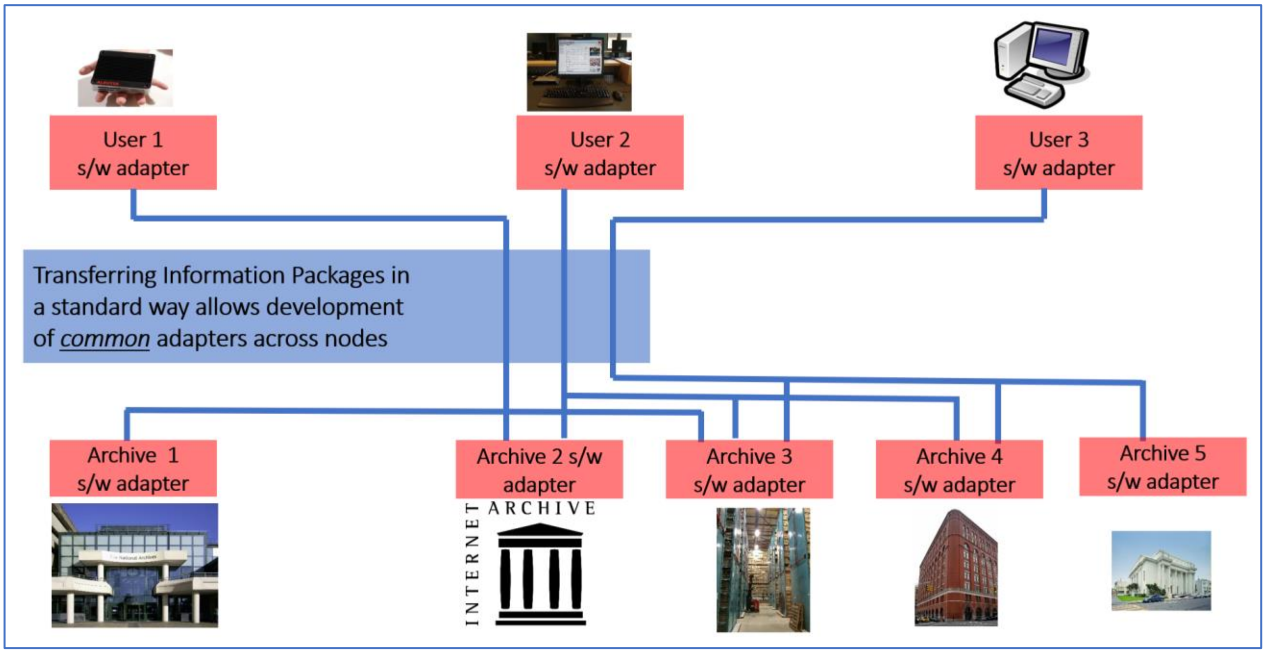

Figure 3: OAIS-IF Application of Adapters

9. OAIS-IF Functional Components

Now to describe the high-level functions of the components of the OAIS-IF Architecture. We will describe the components in Figure 4 going from the top down.

At the top is existing user software which performs functions specific to the user's discipline. This is not part of OAIS-IF, as is the case with the other yellow boxes.

The OAIS-IF User Specific Adapter adapts the User SW interface to the more abstract User Generic Adapter. It may be sensitive to information content that is unique to a user application or a user community (how to handle astronomy images, for example). For OAIS systems, it will probably need to be aware of the structure and content of the transferred Information Packages, including Representation Information and PDI.

The red wavy lines indicate interfaces that will be standardized within OAIS-IF.

The Generic Adaptor on the user side handles the abstract transactions of information objects as they cross the communications interface.

Comm Infrastructure - The Information Packages cross the boundary from user to archive using common communications infrastructure.

The Generic Adapter on the archive side mirrors the functionality of the Generic Adapter on the user side. In fact, the current design should allow for it to be exactly the same adapter software. It will deliver the Information Packages from the comm interface to the Archive Specific Adapter.

The Archive Specific Adapter adapts the Archive's interface to the generic adapters. As with the user specific adapter, to a degree it understands the content of the Information Packages well enough to do common functions, such as repackaging the Archive's response to the User's Data Request. For example, a typical archive may supply only data objects, therefore the Specific Adapter must associate that data object with the appropriate Representation Information, to make an Information Object, and could add various elements of PDI into the Information Package.

The Archive SW is the existing (or legacy) software that hopefully will remain unchanged.

The Switchboard is a resource that describes to a user how to access to any given archive. It may be part of the archive, or it may reside externally, or there may be multiple switchboards (internal and external) that assist users in accessing archives. As a resource for archive users, a switchboard may point to where adapters can be downloaded, for example.

The Registry, like the switchboard, may be internal to the archive, or it may be an external resource. It provides access to additional Representation Information. An external Registry of supplemental metadata may be another archive or simply an archive-like registry run by a designated community. There may be negotiations to allow the user to request more RepInfo, or to have it transformed into an alternate format. This will support the cross-discipline capabilities we discussed earlier.

10. Resources for Developing OAIS-IF Systems

Now to talk about resources available for developing your own OAIS-IF implementation. The DAI team recognizes that cost-effective development is an important criterion for adoption. We have taken measures wherever possible to simplify life for system developers. The layered architecture and the split between the Generic Adapter and Specific Adapter functions is a result of that approach.

For the User Specific Adapters, because of the dependency on the interfaces to the legacy user software, user developers will have to develop most of these. However, we expect that they can be shared across discipline communities (for example, a community like Web Archive users.) Our DAI team in CCSDS will develop a prototype which hopefully can be made open source and available for download as a starting point for discipline-specific developers.

We expect the Generic adapters to be common for many archives and communities. CCSDS plans to produce a prototype of Generic Adapters usable on both the user side and archive side. We hope to make them available to the community, again, as open source so that user and archive developers can capitalize on them.

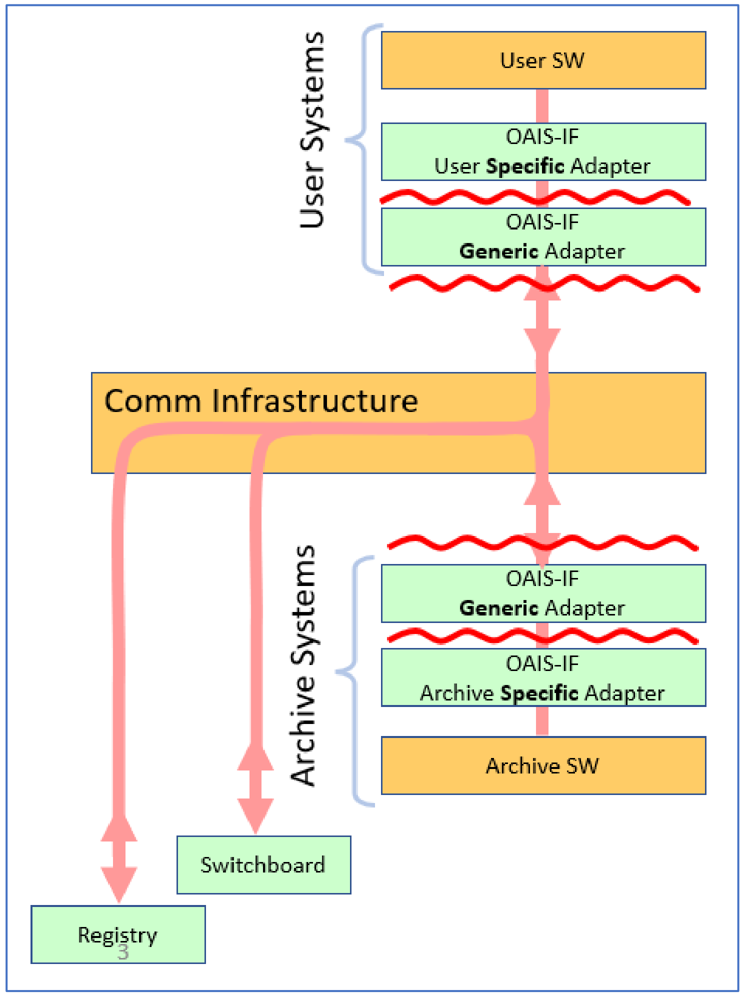

Figure 4: OAIS Functional Components

For the Archive Specific Adapter, archives may need to produce their own software, but (again) we hope to make a CCSDS prototype available as a starting point. And we expect that some communities of archive users (e.g. those that are in a certain discipline or use a certain commercial archive product) can share adapters that they develop, or buy commercial adapters. The best case is that they can use the CCSDS prototype and simply configure it, e.g. to fill in some tables, for example associating the appropriate Representation information with each Data Object.

Both Registry and Switchboard capabilities will be available in the CCSDS prototypes as simple text files. However, they may be developed by archives or a third party at the remote location. Many would consider these functions to be part of any archive's job, but as we have said our architecture accommodates the utilization of remote resources supplied by entities external to the archive. The Registry and Switchboard could be implemented as specialised archives which can then use the same adapters as the archives that they supplement.

11. OAIS-IF Deployment Options

We continue to try to leave as much flexibility as possible in the hands of system developers, without breaking interoperability. See Figure 5 below. The first (left) column shows the baseline deployment that we have been discussing.

The second column, "Consolidated Adapters" shows a possible implementation where the Specific Adapters and Generic Adapters have been merged into one adapter on the user side, and another adapter on the archive side. As long as they comply with interface 1 and include the internal functionality of interface 2, they should still be interoperable with users and archives that use different deployment options.

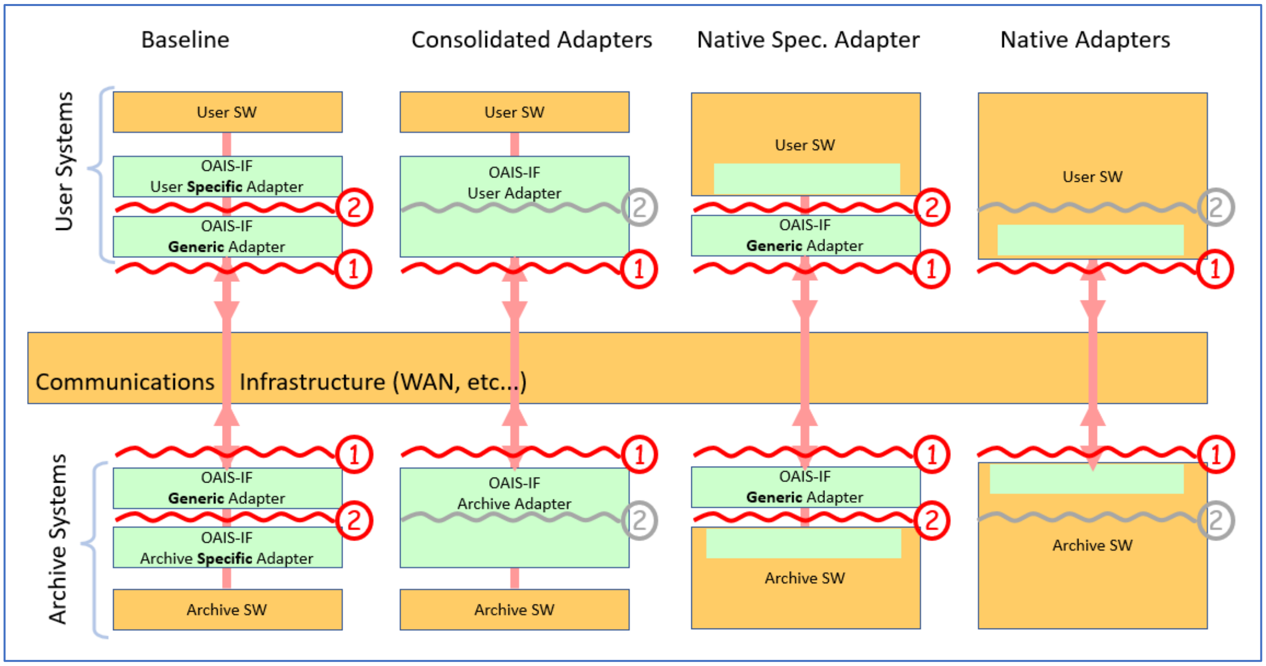

Figure 5: OAIS-IF Deployment Options

The third column, "Native Specific Adapter," shows an option where the Specific Adapter has been merged with the user or archive software. Again, if they comply with interfaces 1 and 2, there should still be no problem with interoperability with external entities.

The fourth column, "Native Adapters" shows an implementation where both of the adapter functions are incorporated with the native user or archive software. Again, still complying with interface 1 for interoperability.

These last three options will lose some of the benefits that come with a layered architecture. However, some development organizations may prefer this more monolithic architecture. The DAI WG always endeavors to allow implementation options for the developers of systems that comply with these interface standards.

12. The DAI WG Document Tree

The documents that the DAI WG plans to produce are indicated in our document tree in 6 below with orange or red indicators. (Already published documents have green status indicators.)

In CCSDS parlance [16], Green books are informative (not normative) books. Magenta books are normative but not necessarily "implementable" or "testable" because of the high level at which the normative requirements are stated. Blue books are normative and implementable/testable. With that in mind…

The division between the OAIS Process Framework (OAIS-PF) documents and the OAIS Interoperability Framework (OAIS-IF) is shown by the dotted line.

The primary work we are concentrating on now is the ADD Magenta Book and the OAIS-IF Core Specifications Blue Book (Generic Adapters, or interfaces 1 and 2 in our earlier figures.)

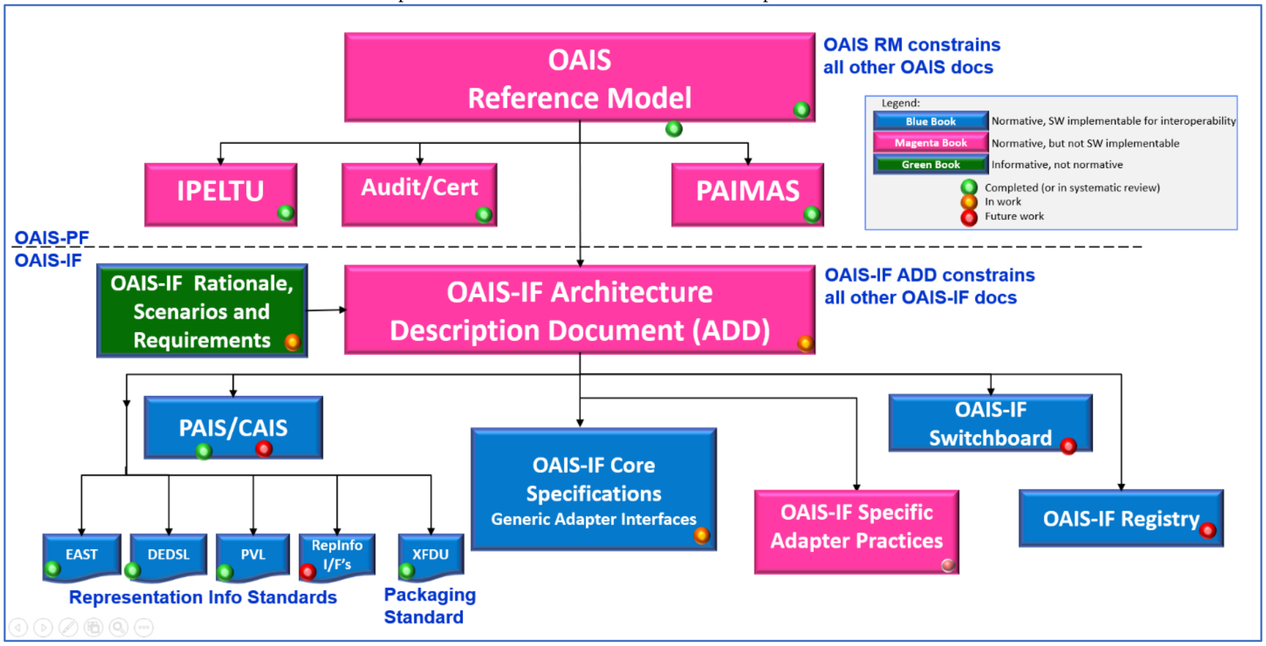

Figure 6: DAI WG Document Tree

13. The OAIS-IF Development Approach

The OAIS-IF Development Approach is shown below in Figure 7. The primary inputs are the Reference Model for OAIS and the various use cases that we have developed or been provided by users or our space agencies. We then develop a UML model of the functions that eventually matures into an architecture. That provides input to the Architecture Description Document. In some cases, the text of the ADD is actually output from the model itself.

As we said earlier, we hope that the model we use to specify the standards and interfaces can be made available, probably on the CCSDS Registry [17], and will be freely downloadable by archive developers to give them a head start on developing systems with OAIS-IF compliance.

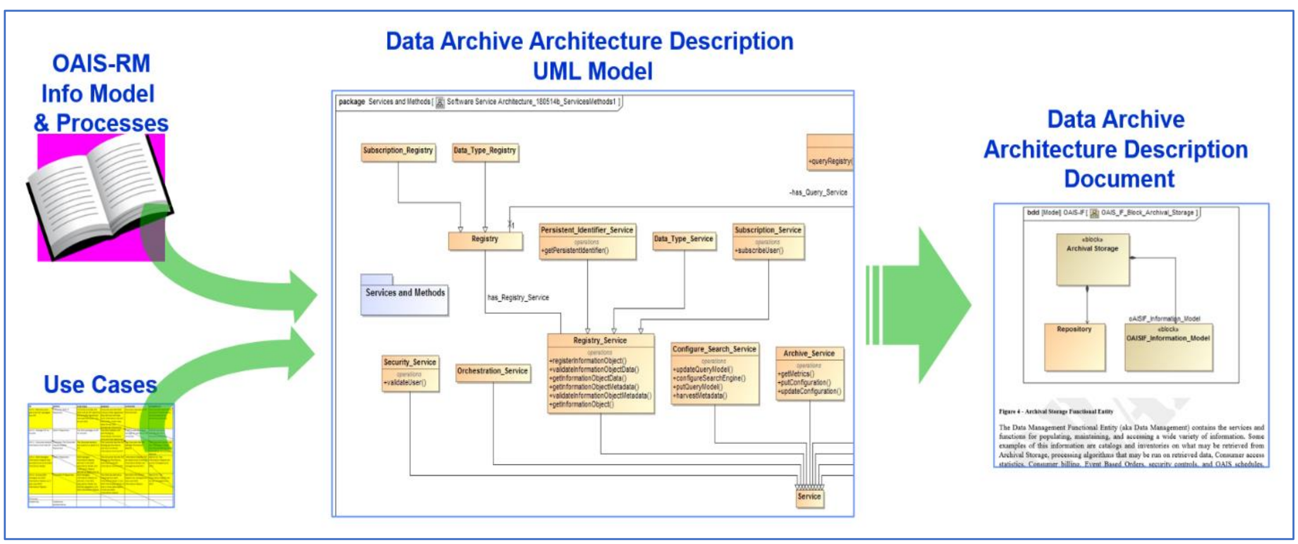

Figure 7: The OAIS-IF Development Approach

14. Conclusions

For all archives, once we have broad availability of generic and specific OAIS-IF adapters, this should result in easy and ubiquitous access to the archive's customers, given the appropriate access rights, of course. The ability to have supplemental access to the specific types of metadata (RepInfo and PDI) made more accessible by supplementary switchboards and registries should dramatically increase the ability of many communities to perform cross-discipline research.

For non-OAIS archives, the same benefits of interoperability with a broader user community can be achieved. Furthermore, the first step of OAIS-IF interoperability can become an evolutionary step towards truly long-term digital preservation with OAIS functions.

For OAIS Archives, the OAIS-IF capability will serve their customers well by giving them rapid access to tools that are designed to service their needs, like Representation Info and Preservation Description Info. As well as the broad interoperability that we described above.

The layered modular architecture approach will allow upgrade and swap-out of adapters as needs and technology change. This, along with the plan for prototypes and resources will enable rapid implementation by the SpaceOps community, and by space mission developers around the globe.

We hope that as the awareness of digital preservation grows in the space community, missions will realize the utility of preserving not only science data but their engineering and design data, to better incorporate lessons learned from past mission developments. And in the upcoming era of AI access to "big data," the ubiquitous access (given access rights, of course) to mission development and science archives will enable AIs to excel with extensive deep learning training material.

The DAI Working Group welcomes review and critique, and participation in the development of OAIS-IF from users or developers that can make technical contributions.

Acknowledgements

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 SpaceOps.

Appendix A: Acronyms and Abbreviations

ADD	Architecture Description Document
AI	Artificial Intelligence
AIP	Archival Information Package
ALSEP	Apollo Lunar Surface Experiment 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)
IF	Interoperability Framework
IMAGE	Magnetopause-to-Aurora Global Exploration
IPELTU	Information Preservation to Enable Long Term Use
ISO	International Standards Organization
ISS	International Space Station
JPL	Jet Propulsion Laboratory
MBE	Model Based Engineering
MOIMS	Mission Operations and Information Management Services
NASA	National Aeronautics and Space Administration
OAIS	Open Archival Information System
OAIS-IF	Open Archival Information System - Interoperability Framework
OAIS-PF	Open Archival Information System – Process Framework
PAIMAS	Producer Archive Interface Methodology Abstract Standard
PAIP	Producer Archive Interface Protocol
PAIS	Producer Archive Interface Specification
PDF	Portable Document Format
PDI	Preservation Description Information
PDS	Planetary Data System
PTAB	Primary Trustworthy Digital Repository Authorisation Body Ltd
RepInfo	Representation Information
RM	Reference Model
SANA	Space Assigned Numbers Authority
SIP	Submission Information Package
SW	Software
UML	Unified Modelling Language
VCS	Voluntary Consensus Standards
WG	Working Group
XFDU	XML Formatted Data Unit

References

[1] CCSDS, About CCSDS, https://public.ccsds.org/about/default.aspx (accessed 11 Jan 2023)

[2] CCSDS, DAI WG, https://cwe.ccsds.org/moims/default.aspx#_MOIMS-DAI (accessed 11 Jan 2023)

[3] CCSDS, CCSDS 650.0-M-2, Reference Model for an Open Archival Information System, https://public.ccsds.org/Pubs/650x0m2.pdf (also known as ISO 14721)

[4] CCSDS, ISO TC20/SC13 Contents, https://public.ccsds.org/about/ISO_TC20-SC13_contents.aspx

[5] ISO, ISO-14721:2012, https://www.iso.org/standard/57284.html

[6] ISO, ISO-16363:2012 Space data and information transfer systems — Audit and certification of trustworthy digital repositories, https://www.iso.org/standard/56510.html, also available as CCSDS 652.0-M-1 Audit and certification of trustworthy digital repositories, https://public.ccsds.org/Pubs/652x0m1.pdf

[7] ISO, ISO-16919:2014 Space data and information transfer systems — Requirements for bodies providing audit and certification of candidate trustworthy digital repositories, https://www.iso.org/standard/56510.html, also available as CCSDS 652.0-M-1, Requirements for bodies providing audit and certification of candidate trustworthy digital repositories https://public.ccsds.org/Pubs/652x1m2.pdf

[8] PTAB Ltd, Primary Trustworthy Digital Repository Authorisation Body Ltd, http://www.iso16363.org/

[9] National Science Foundation, "Cyberinfrastructure Vision for 21st Century Discovery, page 24, https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

[10] NASA GSFC Engineering Colloquia Series – Video of Vint Cerf "Digital Vellum" lecture. https://www.youtube.com/watch?v=HOl4GLjboDw

[11] Kearney, et al, Digital Preservation Archives – A New Future Architecture for Long-term Interoperability, https://arc.aiaa.org/doi/pdf/10.2514/6.2018-2402

[12] Johns Hopkins University Applied Physics Lab, Interstellar Probe - Humanity's Journey to Interstellar Space, https://interstellarprobe.jhuapl.edu/

[13] CCSDS, CCSDS 650.0-M-2, Reference Model for an Open Archival Information System, https://public.ccsds.org/Pubs/650x0m2.pdf, pages 1-10 and 1-14.

[14] CCSDS, CCSDS 650.0-M-2, Reference Model for an Open Archival Information System, https://public.ccsds.org/Pubs/650x0m2.pdf, page 1-12.

[15] CCSDS, CCSDS 650.0-M-2, Reference Model for an Open Archival Information System, https://public.ccsds.org/Pubs/650x0m2.pdf, pages 4-20 to 4-50.

[16] CCSDS, Organization and Processes for the CCSDS, https://public.ccsds.org/Pubs/A02x1y4c2.pdf, page 6-1.

[17] Space Assigned Numbers Authority (SANA) see https://sanaregistry.org/https://sanaregistry.org/