Development of New Standards for OAIS Interoperability

SpaceOps 2023
March 6-10, Dubai UAE

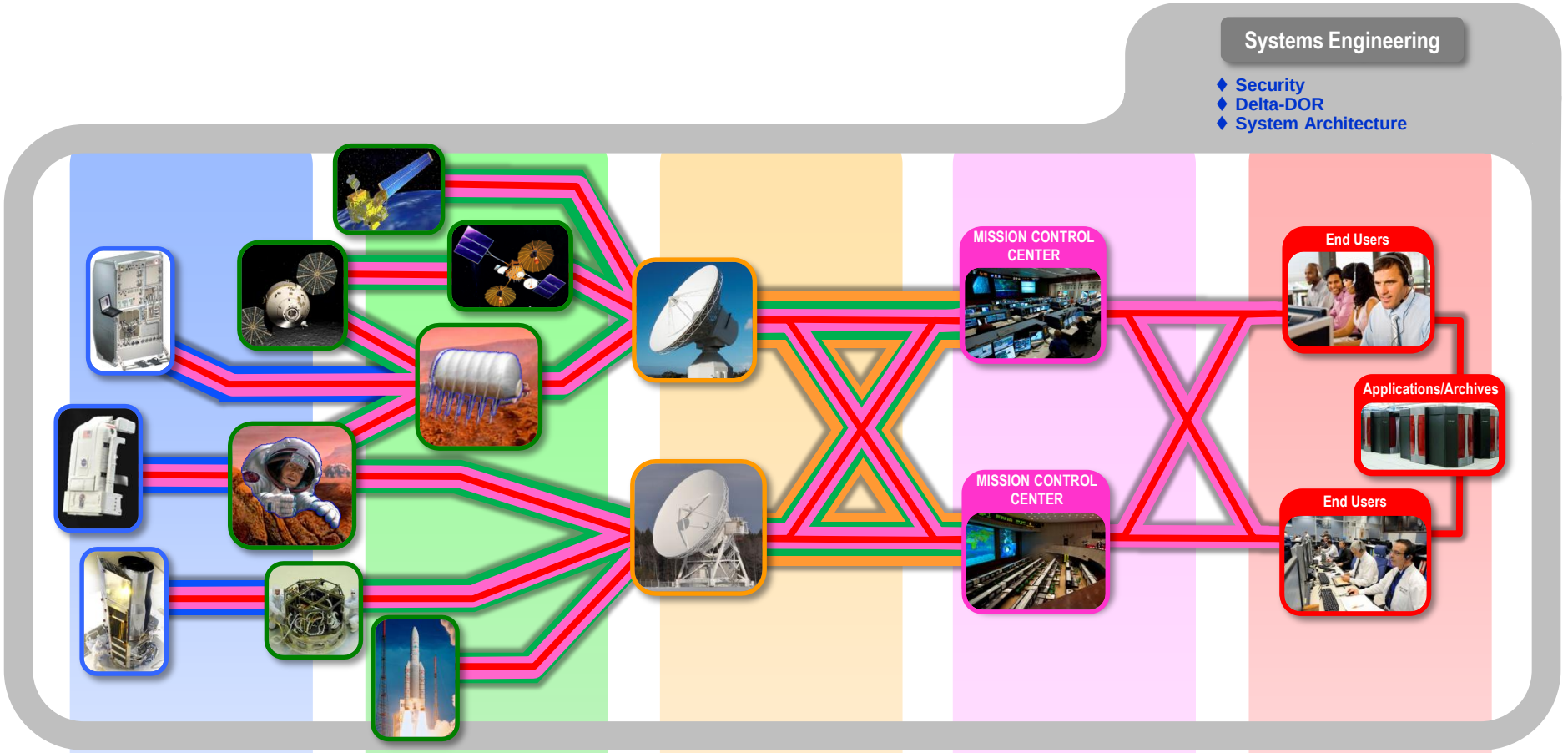
Mike Kearney
KearneySolutions@gmail.com
Space Infrastructure Foundation
Also the CCSDS DAI WG
Also the Google Digital Vellum Project

Supporting Co-authors:
David Giaretta, PTAB Ltd.; DAI WG co-chair
John Garrett, Garrett Software; DAI WG co-chair
Steve Hughes, NASA JPL CalTech; DAI WG Architecture Document Book Editor

Intro to CCSDS and the DAI Working Group

- Reasons for this presentation:
  - Introduce proposed new architecture for archive interoperability
  - Get feedback on our approach, invite participation
  - Elevate awareness of need for long-term digital preservation.
- Background on CCSDS:
  - *Consultative Committee for Space Data Systems*
  - CCSDS is organized by space agencies, but inclusive of other (non-space) agencies, industry and academia
  - About 24 working groups, one of which is the DAI WG
- ***Data Archive Interoperability*** *Working Group*
  - Focused on ***long-term Digital Preservation*** Archives
  - Developed the **Open Archival Info System** (OAIS) ***Reference Model***
  - Published by both CCSDS and ISO, widely adopted for digital preservation

CCSDS Overview - End-to-End Architecture



New Direction for the DAI WG

- ***Initially*** the DAI WG worked on ***processes*** for digital preservation.
  - That resulted in the OAIS Reference Model first published in 2002
  - Continued to optimize OAIS Processes in 5-year reviews.
  - OAIS-RM and associated process documents are now collectively referred to as the ***OAIS Process Framework (OAIS-PF)***
  - Subsequently set up ISO certification process for trustworthy repositories
  - Result - OAIS practices are widely accepted and practiced among national archives and libraries around the world.
- ***Now*** DAI is starting work on ***interoperability*** standards for archives
  - Google's Vint Cerf approached us and advocated stronger focus on technical interoperability for preservation archives.
  - For both OAIS archives and "regular" archives (*your* current archives).
  - Called the ***OAIS Interoperability Framework (OAIS-IF)***.

The Digital Preservation Problem Statement for Space Missions

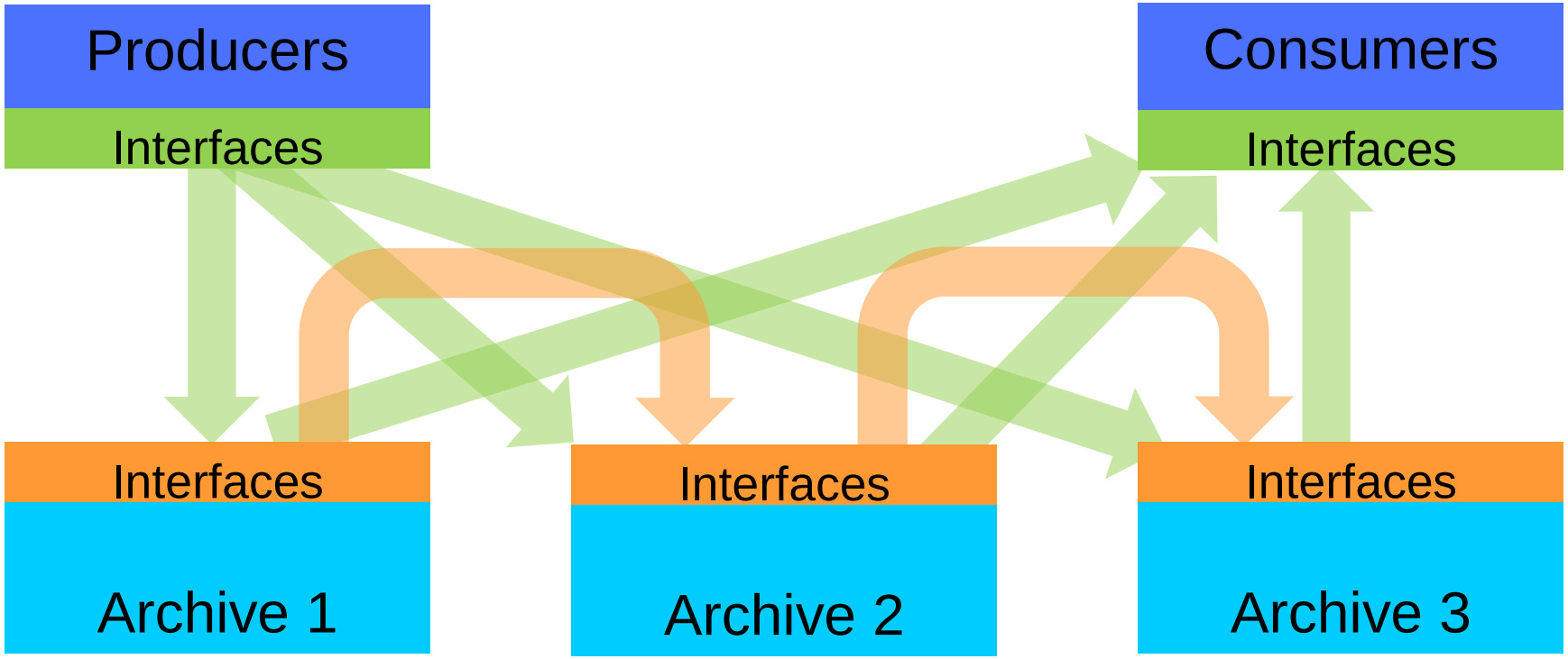
- With a few exceptions, many missions lose access to their mission data after the mission terminates – funding stops – Software becomes obsolete
  - Legacy mission TLM is often valuable but inaccessible.
  - Lost mission and project lessons learned will handicap new programs.
- Lessons learned (or not) examples:
  - Prior SpaceOps presentation examples: Apollo, ALSEP, NASA Galileo, IMAGE.
  - New Use Case: NASA Interstellar Probe mission may last 50+ years.
- Software obsolescence seen as a global problem
  - Much of today's data may be inaccessible in decades, or certainly in a century.
  - Google on "Digital Dark Age"
  - Massive data sets will be needed to train future AIs. Preservation→AI capability.
  - If program/project data is preserved, example query: "For all projects that use cost model X, what percentage of them came in at or under predicted costs?"
  - There is no better time to start to build this capability than right now.

OAIS-IF Development Plans and Constraints

- OAIS-IF builds on OAIS standards (but should work fine with non-OAIS archives.)
- OAIS-IF interfaces can work in parallel with existing archive interfaces.
- OAIS-IF will be made up of a set of interfacing standards, not specifying the underlying functionality of the software, 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 UML models, which should be made available for archive developers building software to the standard.
- CCSDS Requires two independently developed prototypes before a standard can be approved, and we hope that those will be available as references for archive developers, also.

Interoperability for OAIS-IF

- There are many definitions of interoperability
- For our purposes, we plan to provide archival interoperability for:
  1. Users' (Producers' and Consumers') access to archives
  2. Archive-to-archive interoperability (assuming usage of the Users' interfaces)



Current situation

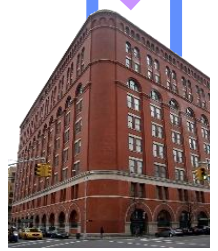
Needs separate software for each type of data

Needs separate software for each type of data

Needs separate software for each type of data

Transferring whatever the archive sends, usually simple files – relies on client having appropriate s/w

INTERNET ARCHIVE



Our vision for OAIS-IF

User 1
s/w adapter

User 2
s/w adapter

User 3
s/w adapter

Transferring Information Packages in a standard way allows development of *common* adapters across nodes

Users always need some “adapters” in order to be able to use new SW without modifying their existing SW.

Archive 1
s/w adapter

Archive 2
s/w adapter

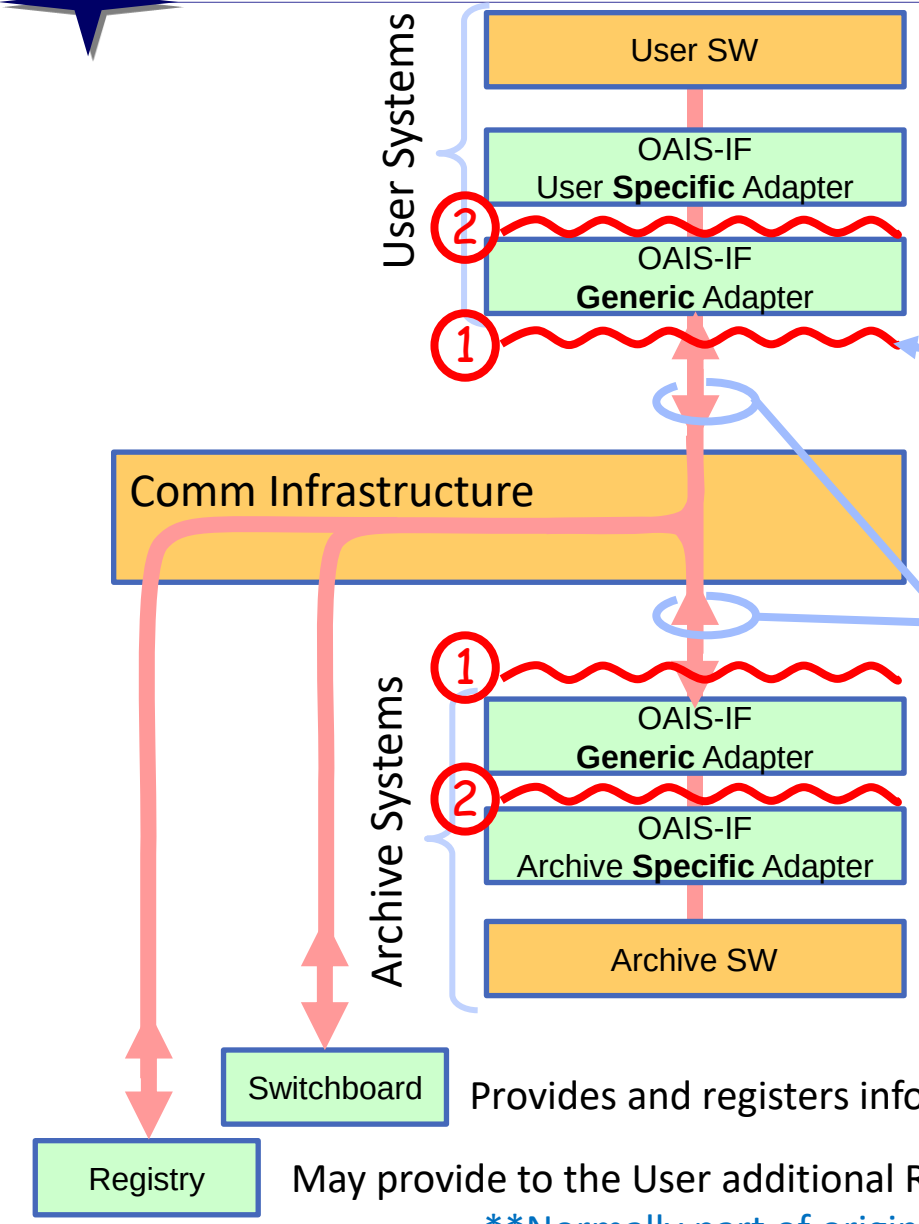
INTERNET ARCHIVE

Archive 3
s/w adapter

Archive 4
s/w adapter

Archive 5
s/w adapter

OAIS-IF Architecture Concept - Functions



Usual functions of existing user software – HMI display, search, etc. (User = producer, consumer or other archives)

Adapts user SW to generic adapter. May have some functions for specific users, driven by the user community.

Handles info packages from archive in abstract fashion – Not sensitive to content of info packages. Meets the standardized interface to the Archive Generic Adapter.

Red waves are interfaces defined by OAIS-IF

Comm Infrastructure performs normal internet communications functions using TCP/IP protocols or other equivalent protocols.

Information Packages flow IAW the OAIS Information Model

Handles Info Packages from User in abstract fashion. Not sensitive to content of info packages. Meets the standardized interface to the User Generic Adapter.

Adapts the Archive legacy interface to the Archive Specific Adapter. May package or repackage Info Objects from the archive to deliver to users.

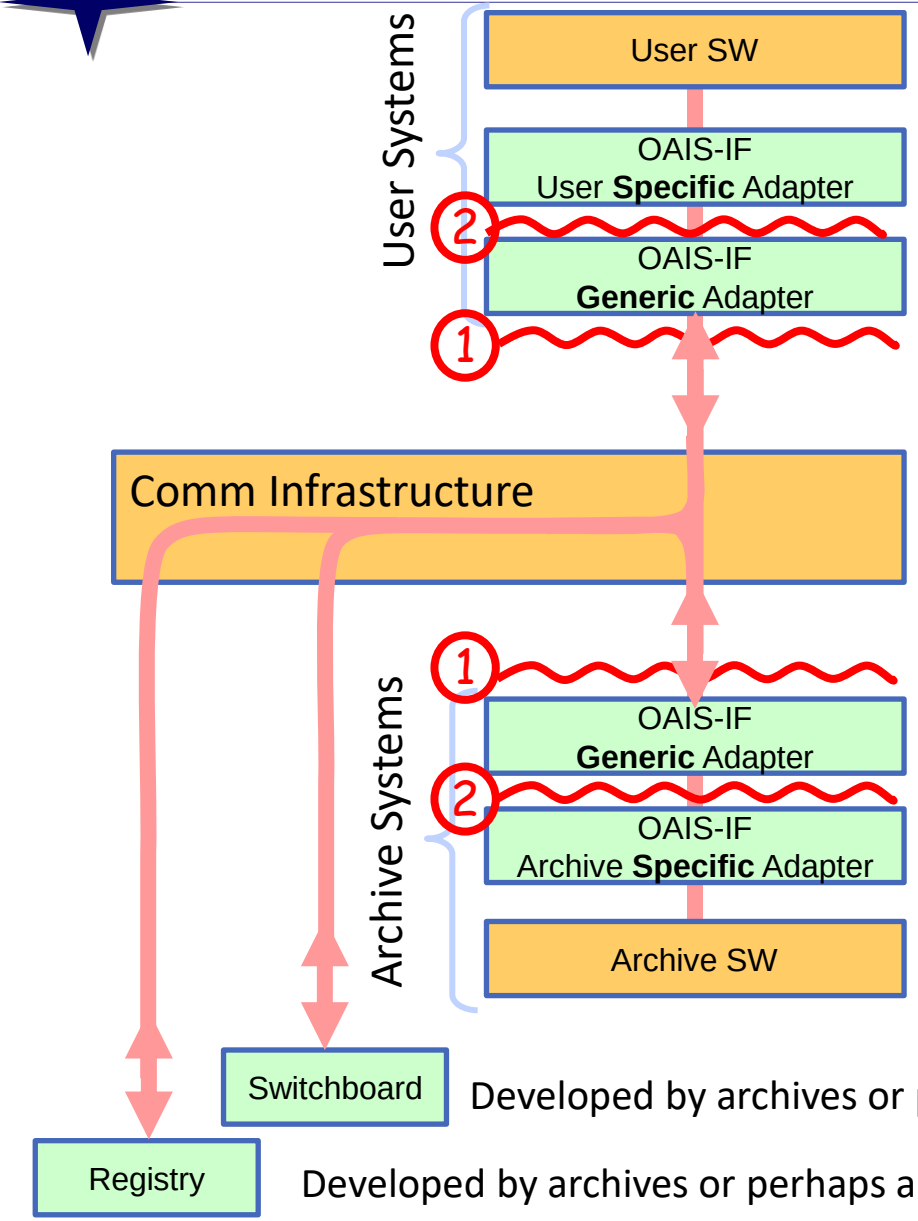
Usual functions of existing archive software – Database responses, retrieval of objects and RepInfo, etc.

Provides and registers info that describes how to talk to an archive.**

May provide to the User additional RepInfo that supplements the RepInfo from the archive.**

**Normally part of originating archive, but may be supplemental data in a remote archive.

OAIS-IF Architecture Concept – Resources



Existing USER software (User = producer, consumer or other archives)

CCSDS will develop a prototype with limited capabilities. User s/w developers will have to write most of these. May be common for a user community.

Prototype available from CCSDS can be used but people may want to write their own

Prototype available from CCSDS can be used but people may want to write their own

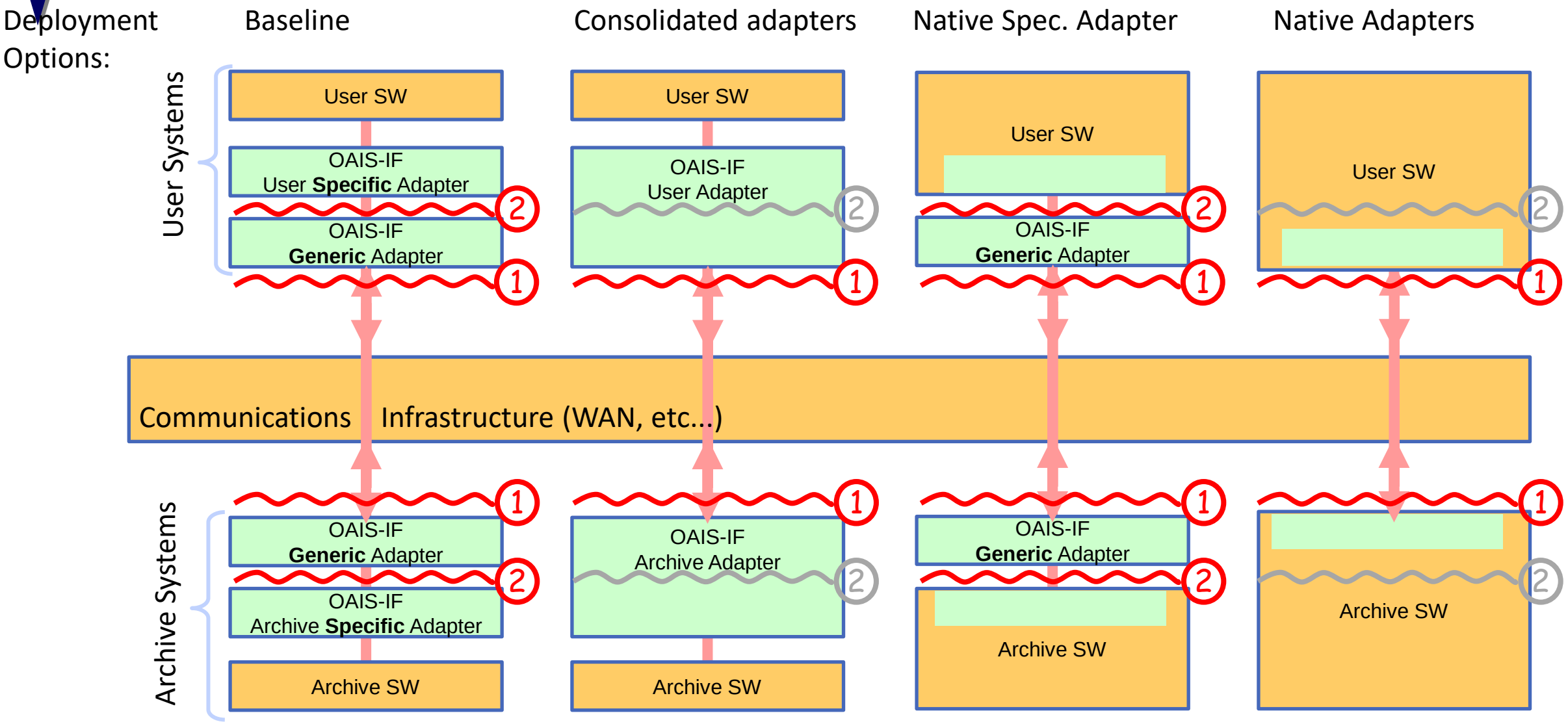
CCSDS hopes to provide a broadly usable implementation which archives would need to configure. Archives may need to produce their own s/w, but we believe that common versions should be sharable between many archive communities.

Existing ARCHIVE software

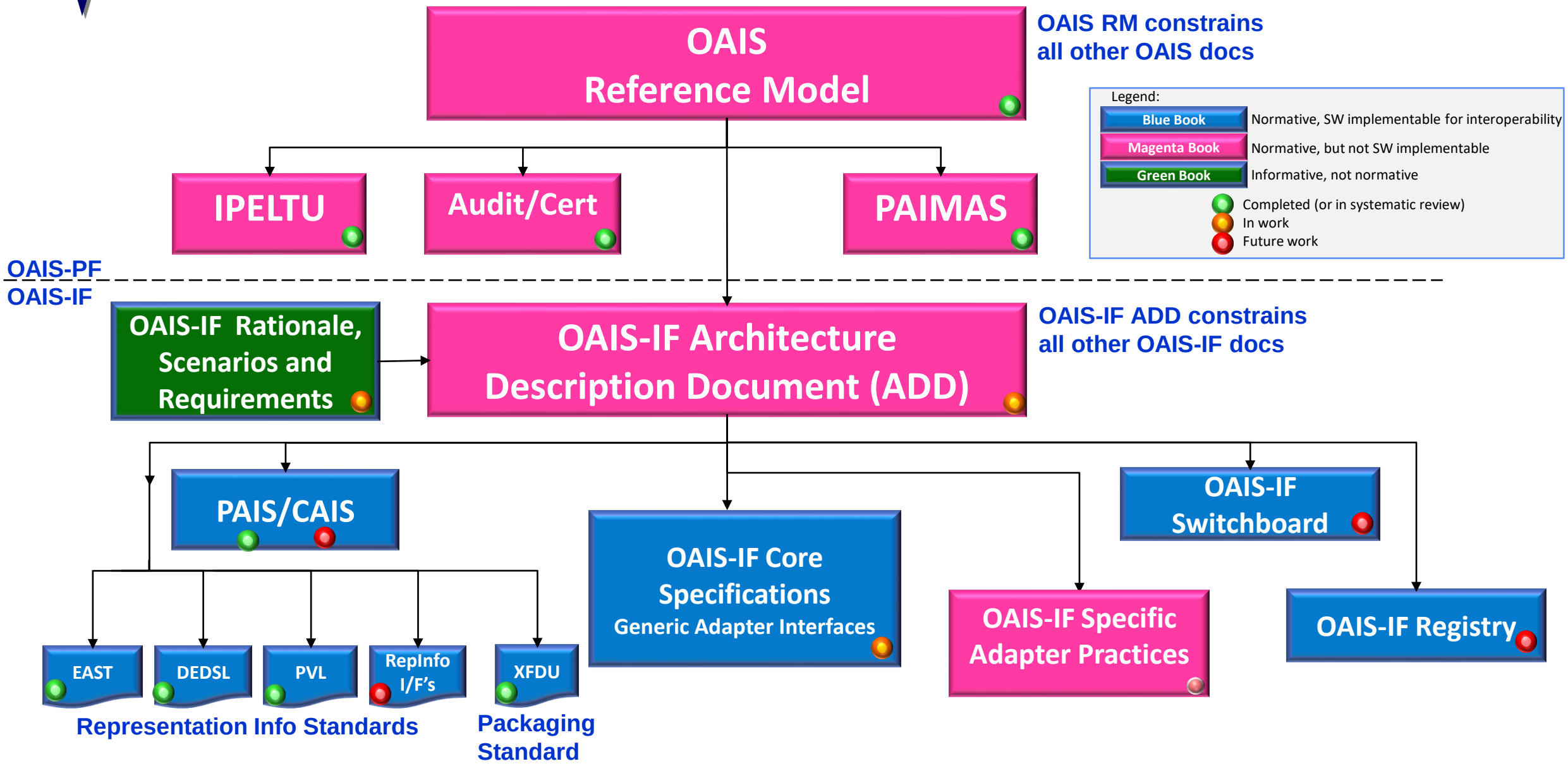
Developed by archives or perhaps a third party source in the designated community

Developed by archives or perhaps a third party source in the designated community

OAIS-IF Architecture Concept – Deployment options



DAI WG Document Tree

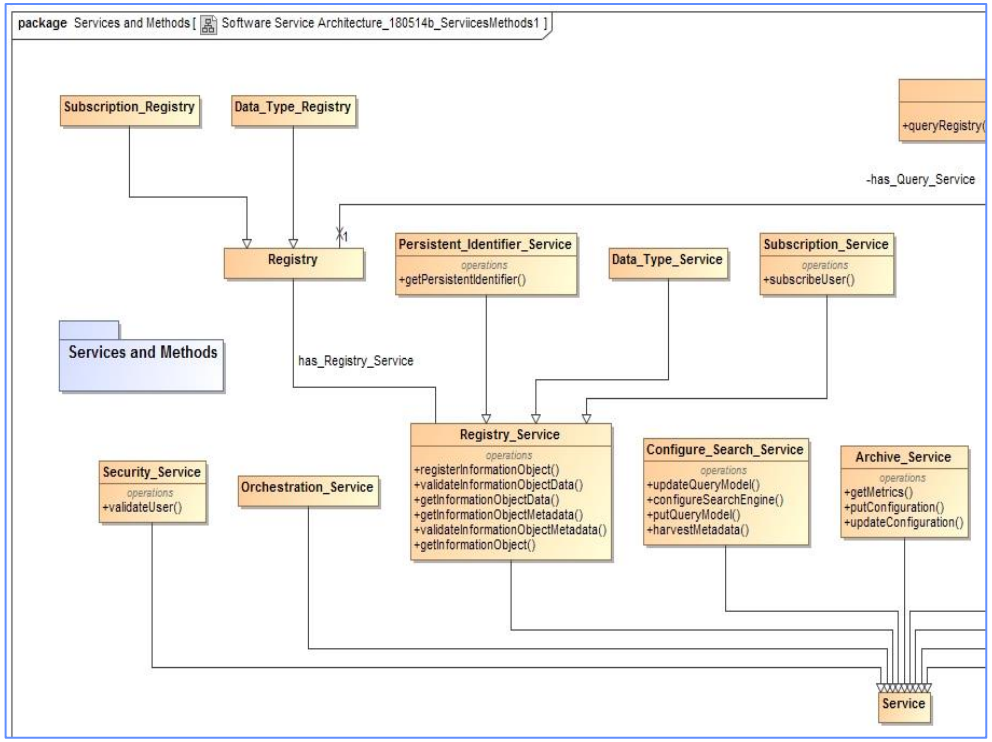


OAIS-IF Development Approach

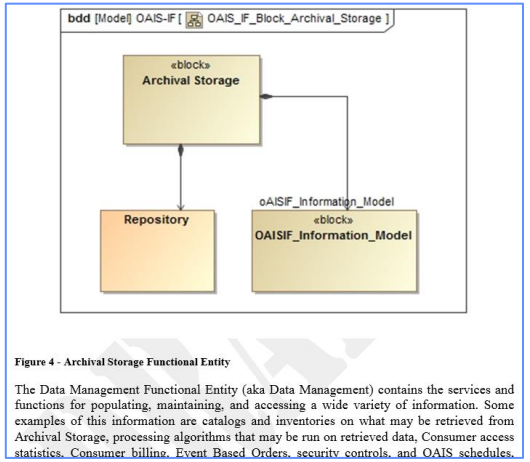
OAIS-RM Info Model & Processes

Use Cases

Data Archive Architecture Description UML Model



Data Archive Architecture Description Document



Key Conclusions

- For all archive types, once we have broad availability of generic and specific adapters, we believe this will spawn greater access between disciplines for cross-discipline research.
- For OAIS Archives, this of course requires Users and Archives to follow OAIS guidelines and store adequate Representation Information (metadata) to allow the interpretation of the preserved data object.
- If *any* archive meets the interface specified by the generic adapter, they should realize the interoperability benefits of OAIS-IF.
- The layered, modular architecture approach will enable rapid implementation by the space operations community.
- By preserving engineering/design data as well as science data, missions will be better able to incorporate lessons learned and techniques from past missions.
- Given appropriate access, this will also foster the “big data” capabilities for AI training with deeply historical data.

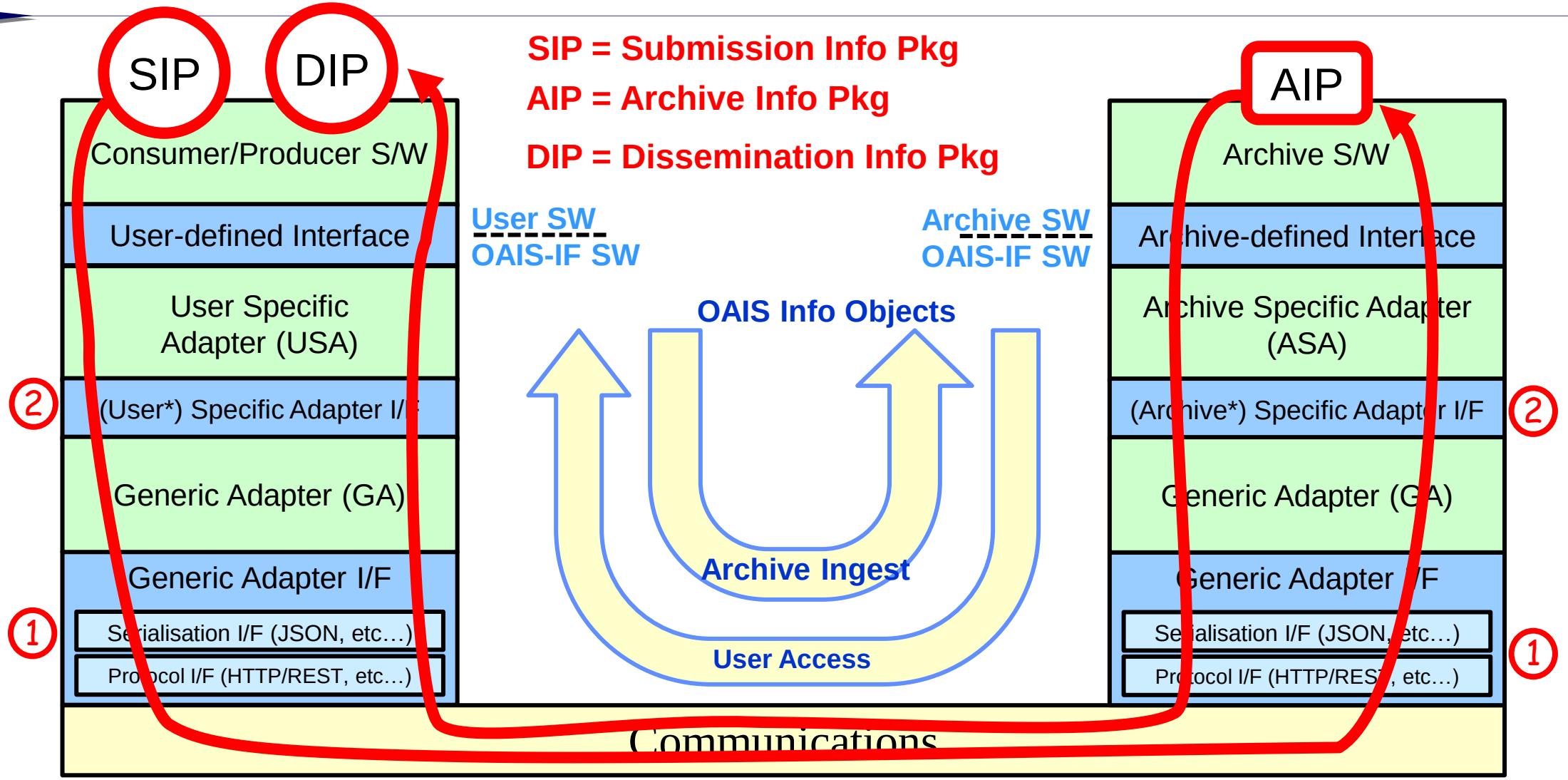
The DAI Working Group welcomes review and critique, and participation in the development of OAIS-IF from users or developers that can make substantive technical contributions.

Official contact info at:
https://cwe.ccsds.org/moims/default.aspx#_MOIMS-DAI

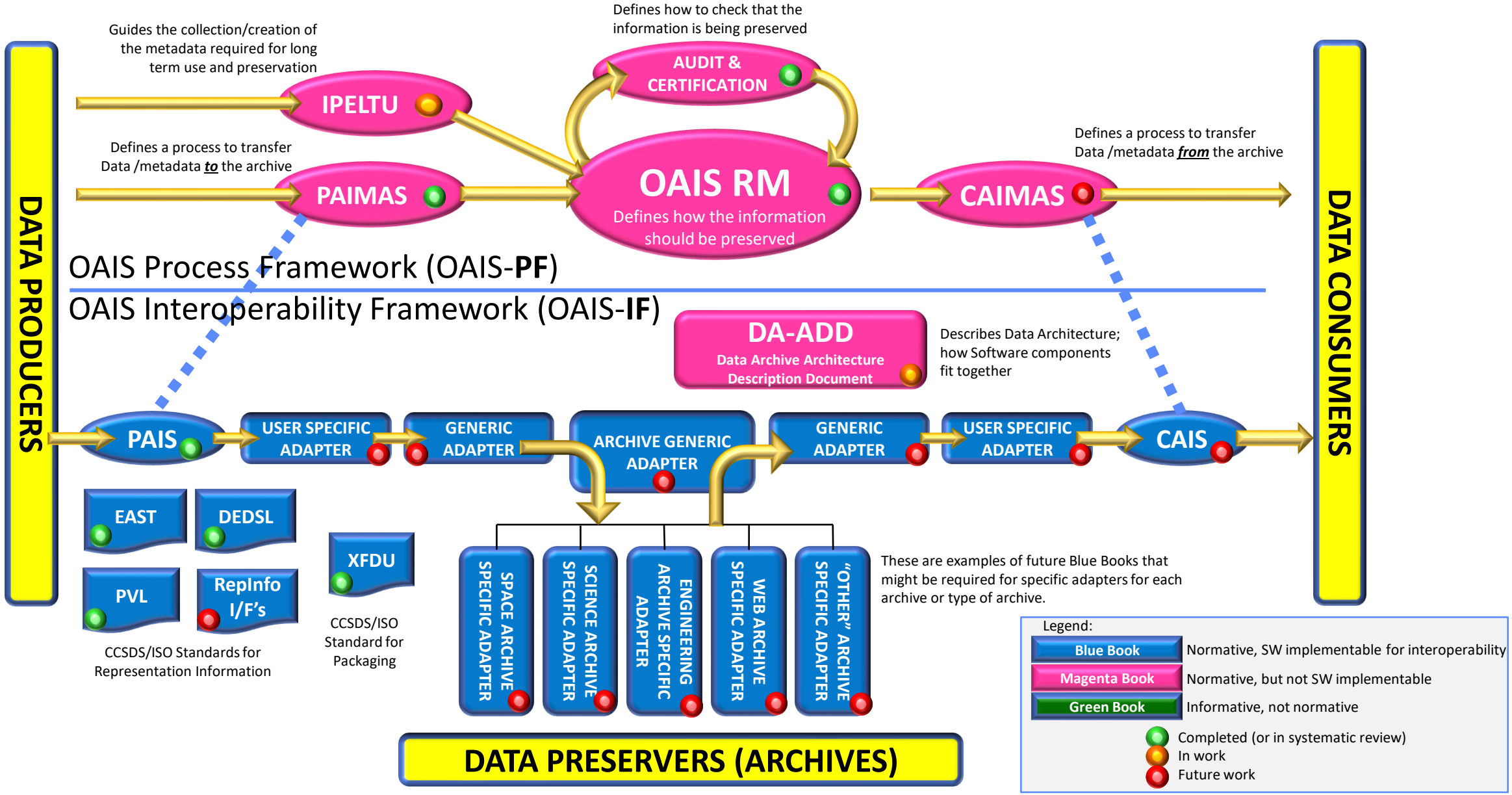
Questions, Comments?

Backup material

Example of OAIS functions

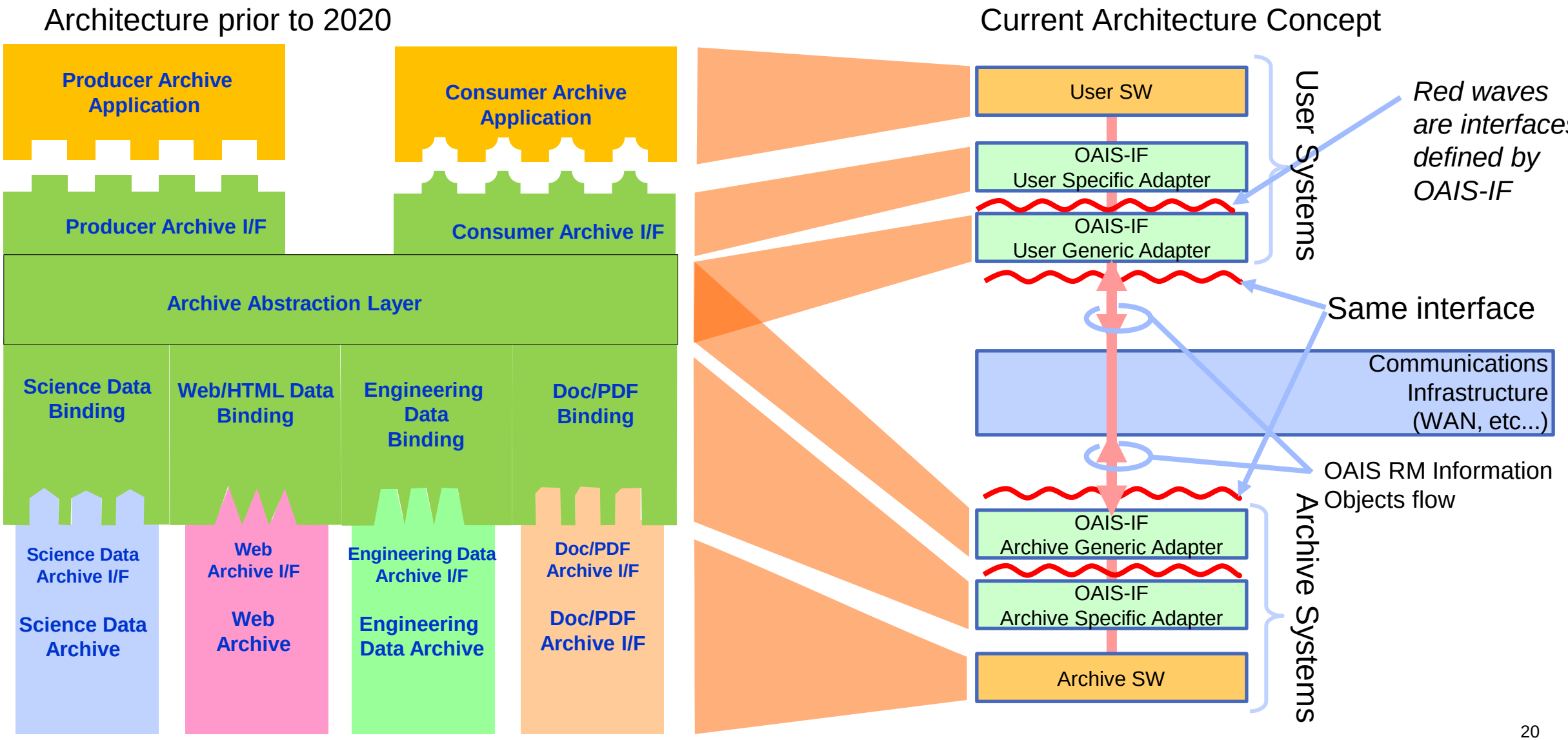


DAI WG Standards and User/Archive Process Flow



Architecture Change Mapping from 2020 to 2021 versions

- For audiences of prior OAIS-IF presentations



Acronym list

Acronym	Description	Link (when available)
ADD	Architecture Description Document	
AI	Artificial Intelligence	
AIP	Archive Information Package	
API	Application Programming Interface	
ASA	Archive Specific Adapter	
CAIMAS	Consumer Archive Interface Methodology Abstract Standard	
CAIP	Consumer Archive Interface Protocol	
CAIS	Consumer Archive Interface Specification	
CCSDS	Consultative Committee for Space Data Systems	www.ccsds.org
DAI WG	Data Archive Interoperability Working Group	https://cwe.ccsds.org/moims/default.aspx#_MOIMS-DAI
DEDSL	Data Entity Dictionary Specification Language	CCSDS 647.1-B-1
DIP	Dissemination Information Package	
EAST	Enhanced Data Subset (Data Description Language)	CCSDS 645.0-G-1
GA	Generic Adapter	
HTML	HyperText Markup Language	
I/F	Interface	
IPELTU	Information Preservation to Enable Long Term Use	
ISO	International Organization for Standardization	www.iso.org
JPL	Jet Propulsion Laboratory	www.jpl.nasa.gov
NASA	National Aeronautics and Space Administration	www.nasa.gov
OAIS	Open Archival Information System	https://public.ccsds.org/Pubs/650x0m2.pdf
OAIS-IF	OAIS Implementation Framework	
OAIS-PF	OAIS Process Framework	
OAIS-RM	OAIS Reference Model	CCSDS 650.0-M-2
PAIMAS	Producer Archive Interface Methodology Abstract Standard	CCSDS 651.0-M-1
PAIP	Producer Archive Interface Protocol	
PAIS	Producer Archive Interface Specification	CCSDS 651.1-B-1
PDF	Portable Document Format	
PVL	Parameter Value Language	CCSDS 641.0-B-2
SANA	(CCSDS) Space Assigned Numbers Authority	www.sanaregistry.org
SIP	Submission Information Package	
SW	Software	
UML	Unified Modeling Language	https://www.omg.org/spec/UML/About-UML/
USA	User Specific Adapter	
WAN	Wide Area Network	
XFDU	XML Formatted Data Unit	CCSDS 661.0-B-1