Ensuring the Ongoing Adoption and Use of a Preservation Content Standard for NASA's Earth Observation Data

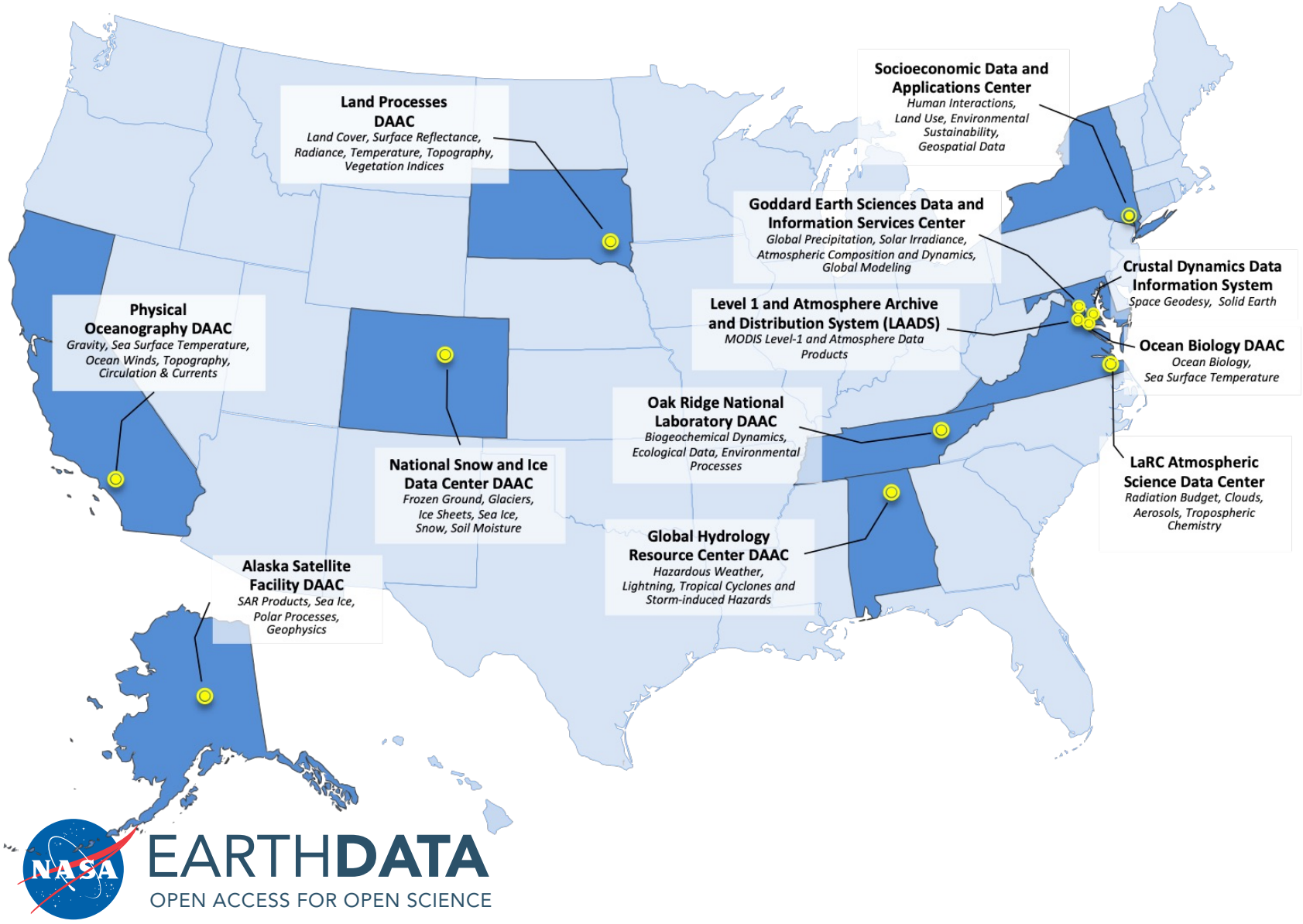
Francis Lindsay, ESDIS Project, NASA
John Moses, SSAI, and NASA ESDIS Project
Hampapuram Ramapriyan, SSAI and NASA ESDIS Project
Jeanne Behnke, NASA ESDIS, Emeritus

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Overview

- The thoughtful and deliberate preservation of data and associated knowledge (provenance and context) ensures that future **users can understand and fully utilize** NASA's expanding archive of scientific data.
- For more than a decade, NASA has been adding its **Preservation Content Specification** (PCS) as a **requirement for new missions** and seeking compliance on a "best efforts" basis for older missions, those that were initiated prior to 2012.
- Many satellite **missions have gathered preservation artifacts** called for in the PCS and have helped NASA further down the path of realizing the full benefits of the PCS.
- The PCS now includes a **mapping to and from ISO 19165-2** published in 2020, which NASA took a lead in developing.
- More recently, in 2022 the NASA PCS team published a revised version of PCS including specifications for airborne and field campaign data and developed a **Preservation Content Implementation Guide** (PCIG).
- This paper seeks to provide readers a current look into the NASA's PV-related activities and examines the challenges that remain for further PCS implementation across the Earth sciences.

EOSDIS Distributed Active Archive Centers (DAACs)



- Discipline-based active and long-term archives of NASA Earth science data
- Serving a large and diverse user community by providing capabilities to search and access science data products and specialized services since 1994
- Continuously evolving to keep up with technology and changing user needs and expectations
- Recent adoption of cloud-based storage and computing capabilities further expands usefulness and potential of EOSDIS

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Preservation Content Specification Categories

Organizes critical pieces of Earth remote sensing data provenance

1. **Measuring Instrument & Platform Description:** Instrument and platform specifications, pre-flight or pre-operational characteristics, performance measurements, calibration method, radiometric & spectral response, noise characteristics, detector offsets, calibration data

2. **Products (Data):** Raw instrument and platform data, derived data products and metadata

3. **Product Documentation:** Structure and format with definitions of all parameters and metadata fields; algorithm theoretical basis; processing history and version history; quality information.

4. **Instrument Calibration:** Instrument/sensor and platform calibration methods (in operation), lookup tables and data; instrument and platform event and maneuver data

5. **Science Algorithm Software:** Science data product generation software and documentation

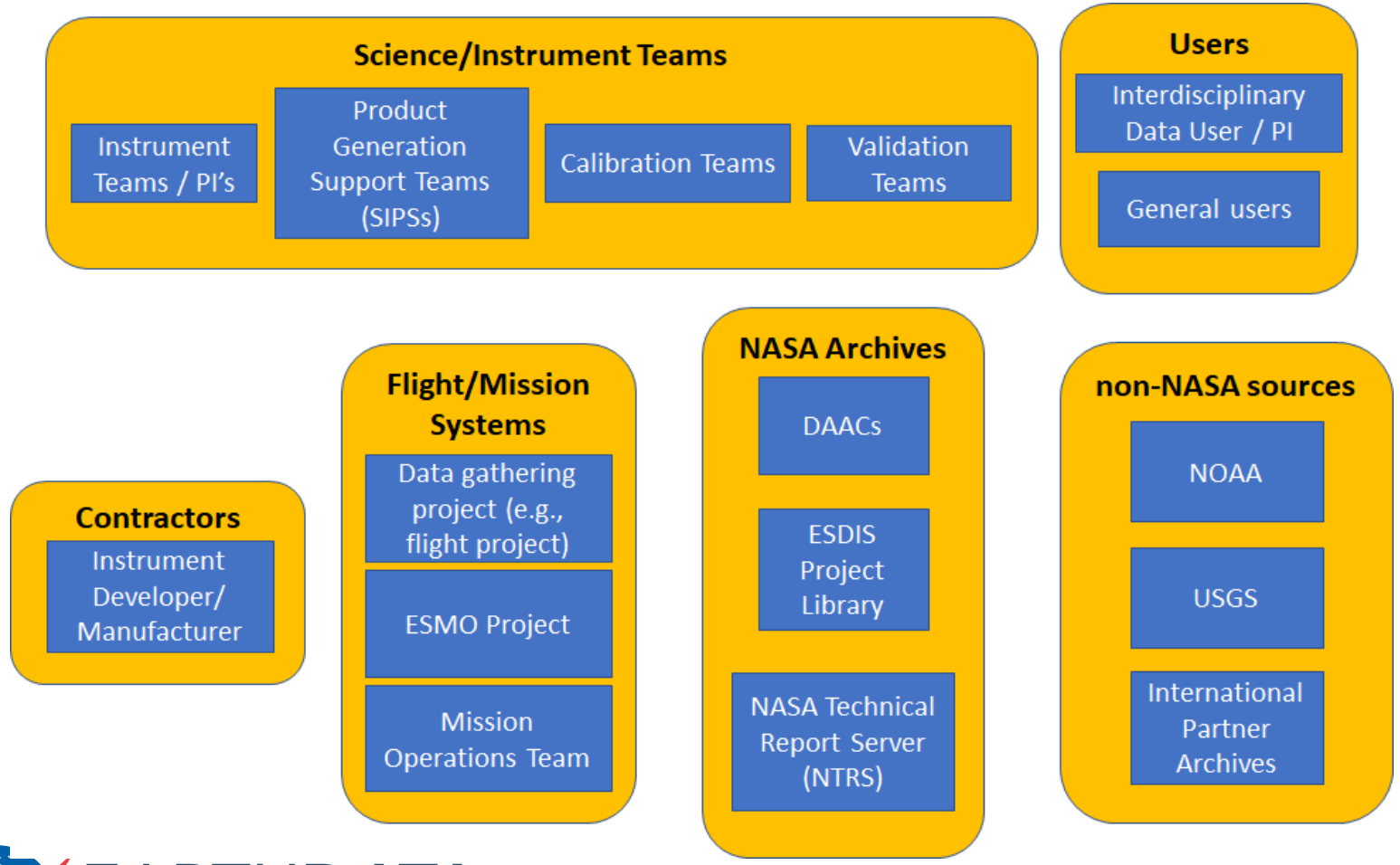
6. **Science Data Product Algorithm Inputs:** Ancillary data or other data sets used in generation or calibration of the data or derived product, description, and documentation.

7. **Science Data Product Validation:** Datasets and documentation describing product accuracy.

8. **Science Data Access and Analysis Tools:** source code or tools for raw, ancillary, product data.

Preservation Content Implementation Guidance

- On orbit/satellite missions



- Figure shows organizations holding preservation contents during the life of a project. These contents must be gathered from these organizations for preservation
- Guidance addresses each section of PCS and indicates how artifacts called for are to be gathered and organized for preservation

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Preservation Content Implementation Guidance

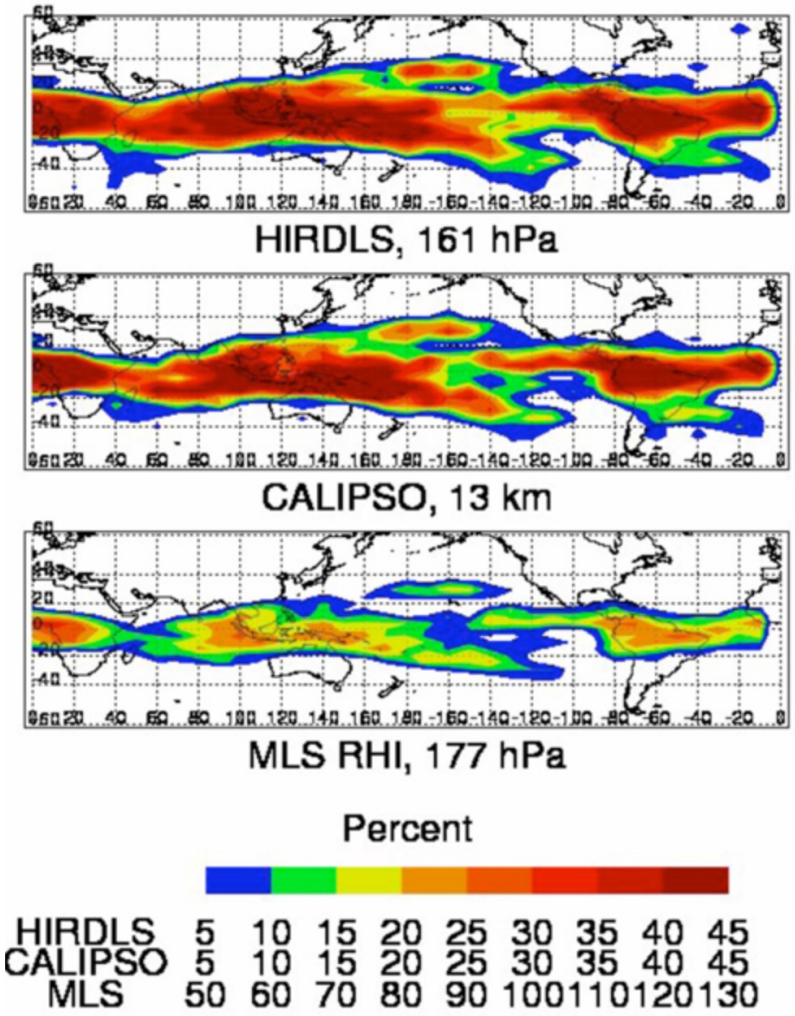
- Recognizes diversity of size, funding, and requirements of NASA-funded projects
  - Expected compliance is indicated as High, Medium or Low depending on project type (e.g., on-orbit missions, research and analysis investigations, data production investigation such as MEaSUREs)
  - Each item called for in PCS is categorized as Required, Suggested, or Not applicable
  - Detailed guidelines are provided in several sections
- Common guidelines applicable to all types of projects
  - Identifying preservation items
  - Responsibility of organizations for identifying and gathering preservation items
  - Provision of pointers to locations of items when not all of them are located at a given DAAC
  - Use of persistent identifiers
  - Ensuring future usability and need for on-going updates
  - What to do when some artifacts are unavailable
  - Organization and presentation of preservation checklists
  - Presentation of preservation items on DAAC websites and project websites
  - Long-term care and maintenance of preservation checklist and artifacts

Preservation Content Implementation Guidance

- Airborne and Field Investigations
  - Guidance identifies types of documents that may apply to various PCS sections
  - Tables show how items unique to airborne and field investigations map to PCS
  - In presenting preservation materials, DAACs may choose whatever manner provides the best clarity for a given investigation. For example:
    - By instrument (and resulting data products)
    - By platform (and the instrument data and derived products associated with that platform
    - By data product type (such as grouping the various aerosol data products, meteorological products, and support data products).
  - Unique cases considered: Non-public Low-Level Data, Extremely Large Investigations, Data at Multiple DAACs, Facility Instruments contributing to multiple focus areas and investigations

PCS Example: High Resolution Dynamics Limb Sounder (HIRDLS)

- Flown on-board NASA's Aura satellite
- Scanning infrared limb-scanner radiometer
- Designed to sound upper troposphere, stratosphere, and mesosphere to determine temperature, concentrations of chemical components of the atmosphere and aerosols along with the locations of polar stratospheric clouds and cloud tops.
- Science team objectives:
  - provide sounding observations with horizontal and vertical resolution superior to that previously obtained
  - observe lower stratosphere with improved sensitivity and accuracy; and
  - improve understanding of atmospheric processes through data analysis, diagnostics, and use of two- and three-dimensional models.
- HIRDLS stopped collecting data on March 17, 2008.



https://aura.gsfc.nasa.gov/science/feature-102009g.html

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

DAACs have approached using the PCS when an instrument ceases operation - HIRDLS Example

- Following stopping observations, the HIRDLS Science Team set to work to finish their final reprocessing (version 7) of all observation products by December 2012.
- As planned these data were delivered to the Goddard Earth Sciences Data and Information Service Center (GES DISC) DAAC for data access and archive.
- Following delivery of data, Science Team funding was scheduled to be phased out.
- Fortunately, two years earlier, a HIRDLS preservation Working Group was formed with participation from US, UK Science Team, and British Atmospheric Data Center representatives (a partner archive), the DAAC and SIPS leaders
    - Working Group came together to review the fulfilling content in the new specification.
    - HIRDLS Science Team helped organize the artifacts, accurately describe the content and critically, how these relate (or map) to the PCS provenance categories.
    - One of these coordination activities resulted in mission calibration data developed by UK Science Team (Oxford) being shared with the US Science Team so that a copy could be archived with the HIRDLS data products.

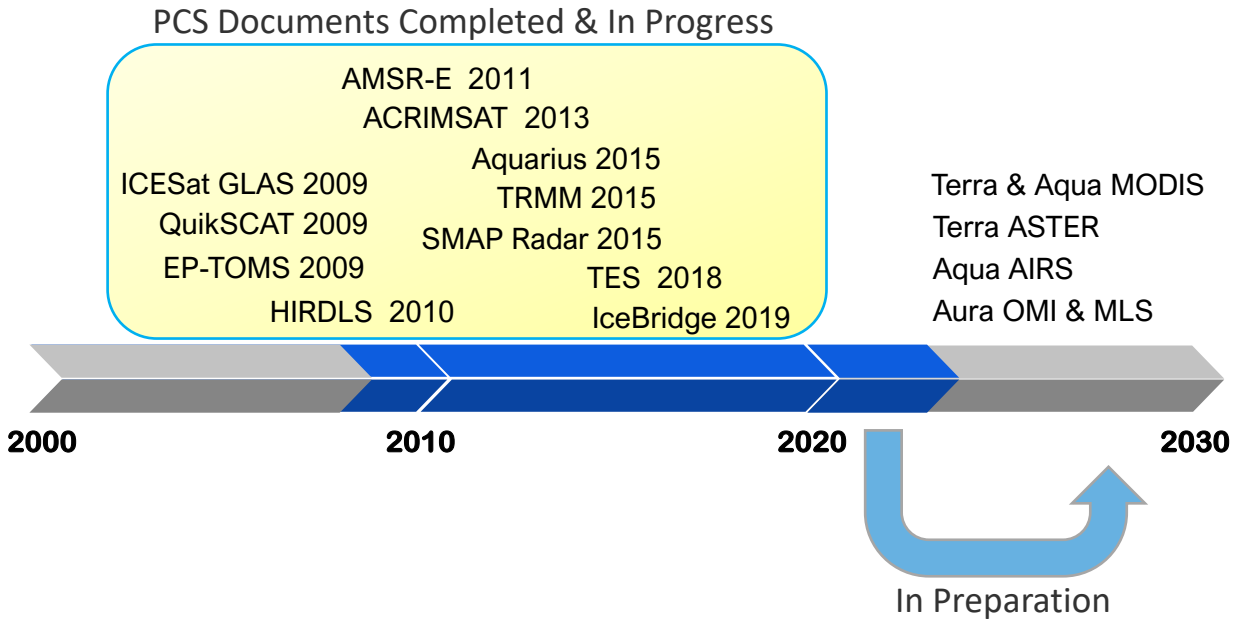
HIRDLS Preservation Results

PCS Category		No. Items	Submitted by	Finish Date
Measuring instrument & Platform Description	HIRDLS Instrument level requirements and specifications, as-built configurations, instrument-level verification, PDR, CDR	214 (+500 drawings)	NCAR Science Team and Oxford University	May 31, 2013
Raw and Derived Products	HIRDLS Calibration Geolocated Radiances & Corrections and Geophysical Parameters. Data products at GES DISC.	34 packages*	NCAR	April 10, 2013
Product Documentation	Product team, requirements, design, process and algorithm versions, product generation algorithms, quality descriptions and pubs	19	NCAR	June 26, 2013
Mission Calibration	HIRDLS Calibration method and data	11	NCAR	April 11, 2013
Science Algorithm	Source Code, versions & documentation	26	NCAR	June 3, 2013
Algorithm inputs	Effect of Using CLCMLS Data in Retrievals	1	NCAR	April 29, 2013
Product Validation	Validation plan and papers (citation list)	2	NCAR	June 27, 2013
Access & Analysis	Orbit Plot - Source Code and description	1 package	NCAR	April 26, 2013

* A package reflects/summarizes the contents of a HIRDLS sheet comprised of multiple items in various formats.

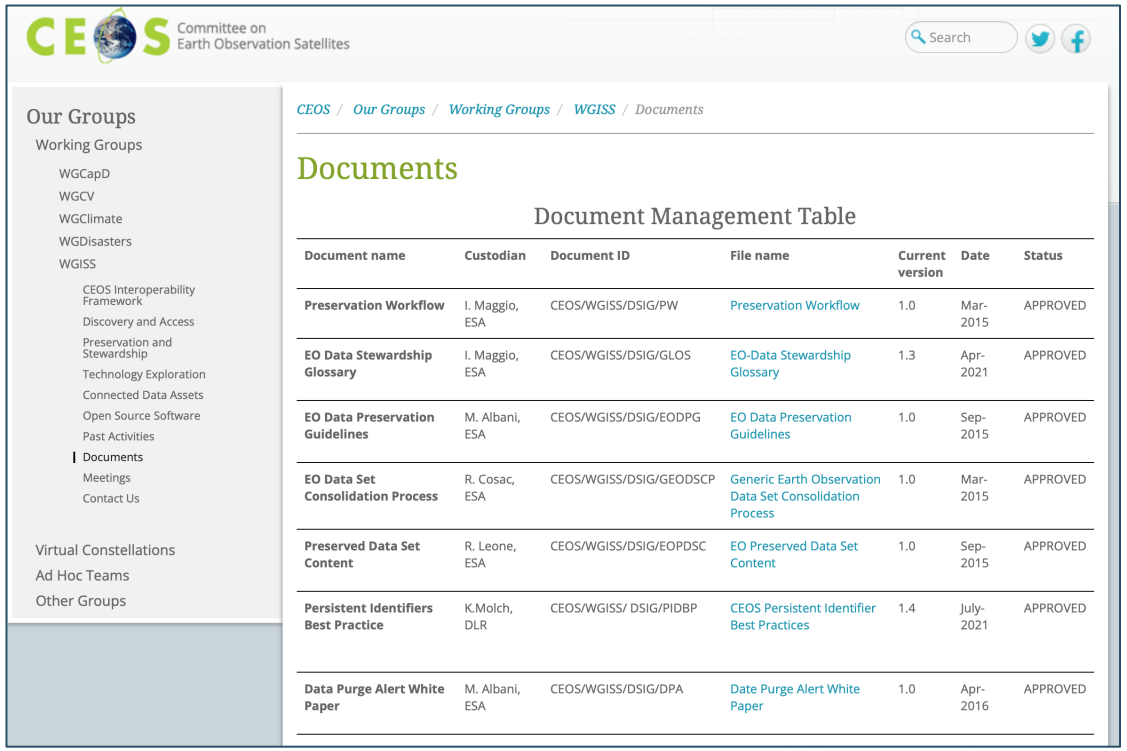
Reaching Mission End-of-Life

- Nothing lasts forever – preservation becomes critical as instruments/missions approach end-of-life
- Number of major NASA missions that have ended instrument observation collections has doubled over the previous ten-year period
- Recently EOS Flight Systems announced platform decommissioning plans that will need to occur over the next few years
    - End-of-life of EOS Terra, Aqua, Aura satellite observations



EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

NASA Participates in Preservation Activities



Document name	Custodian	Document ID	File name	Current version	Date	Status
Preservation Workflow	I. Maggio, ESA	CEOS/WGISS/DSIG/PW	Preservation Workflow	1.0	Mar-2015	APPROVED
EO Data Stewardship Glossary	I. Maggio, ESA	CEOS/WGISS/DSIG/GLOS	EO-Data Stewardship Glossary	1.3	Apr-2021	APPROVED
EO Data Preservation Guidelines	M. Albani, ESA	CEOS/WGISS/DSIG/EODPG	EO Data Preservation Guidelines	1.0	Sep-2015	APPROVED
EO Data Set Consolidation Process	R. Cosac, ESA	CEOS/WGISS/DSIG/GEODSCP	Generic Earth Observation Data Set Consolidation Process	1.0	Mar-2015	APPROVED
Preserved Data Set Content	R. Leone, ESA	CEOS/WGISS/DSIG/EOPDSC	EO Preserved Data Set Content	1.0	Sep-2015	APPROVED
Persistent Identifiers Best Practice	K.Molch, DLR	CEOS/WGISS/ DSIG/PIDBP	CEOS Persistent Identifier Best Practices	1.4	July-2021	APPROVED
Data Purge Alert White Paper	M. Albani, ESA	CEOS/WGISS/DSIG/DPA	Date Purge Alert White Paper	1.0	Apr-2016	APPROVED

- NASA participates in conferences and working groups with NOAA, USGS, other US Earth Science organizations such as ESIP, the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS) and international geographic and remote sensor information standards organizations to discuss and coordinate development of best practices and standards for Earth science data and information.
- Discussions with ESA and CEOS WGISS have been very helpful in development of ISO 19165-2 merging contents of NASA's PCS and ESA/WGISS LTDP – especially, to cover "what" and "when" aspects of collecting preservation content
- **WGISS-55 Meeting** held during April 18-20, 2023
  - Included discussion of a new Interoperability Framework
  - Ensured inclusion of Preservation as one of several "Factors" that must work together to improve interoperability

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

SPD-41a: Scientific Information Policy for the Science Mission Directorate (2022)

- Provides guidance on open sharing of publications, data, and software created in the pursuit of scientific knowledge.
- Builds upon the core principles of openness, equity, and security for SMD-funded research.
- Requires that:
  - Publications be made openly available
  - Research data and software be shared at the time of publication or at the end of funding award
  - Mission data be released as soon as possible
  - Unrestricted mission software be developed openly
  - Science workshops and meetings be held openly to enable broad participation.
- Removes most barriers to fulfilling the preservation content specification for new NASA funded missions and serves as motivation for fulfilling the specification for existing active missions to the extent possible before the end of funding.

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Welcome to A Year of Open Science!

NASA, along with the White House and other federal agencies, has declared 2023 A Year of Open Science to celebrate the benefits and successes created through the open sharing of data, information, and knowledge. Open science is the foundation of NASA Earth Science Data Systems (ESDS) Program activities and accelerates scientific research and understanding.

A special emphasis will be made throughout 2023 to instill open science values and ethics to a wider community of data users and broaden the use of these data. NASA's Transform to Open Science (TOPS) mission has many open activities, events, funding opportunities, and trainings planned over 2023.

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

https://www.earthdata.nasa.gov/news/year-of-open-science

A Few Conclusions - Looking Ahead

- Mission and instrument complexity: near-term and future Earth science missions now being developed and planned often will share a common trend in the increase of complexity in the instruments and anticipated data and the resulting data structures to be used in analysis and application.
- User expectations for access to preservation information: the growth in the number of upcoming observations is another critical aspect of why NASA will need help from our user communities in addressing the preservation issue. Consistently with NASA's policies of opening our data, software, and services to all those who wish it, we'll need to address the users who will expect adequate preservation information is included and is conveniently accssible.
- Leveraging data system technologies to aid in preservation tasks: Our current efforts in the logical collection and organization of preservation content has shown this is no small task and the implementation of collecting, accessing, and archiving of such data is likely to be daunting.
- Timeliness in identifying and collecting preservation content: There will be significant savings in effort if projects or missions plan early in their life cycle to identify and gather the preservation artifacts, rather than wait until the end of life.

References and Resources

CCSDS, 2012: Reference Model for an Open Archival Information System (OAIS). Recommended Practice, CCSDS 650.0-M-2. Magenta Book. Issue 2. Washington, D.C.: CCSDS, June 2012. [Equivalent to ISO 14721:2003] http://public.ccsds.org/publications/archive/650x0m2.pdf

ESA and CEOS/WGISS 2015: Earth Observation Preserved Data Set Content (PDSC), http://ceos .org/ document management/ Working Groups/ WGISS/ Interest Groups/ Data_Stewardship/Recommendations/ EO %20Preserved %20Data %20Set %20Content v1. 0. pdf

ISO, 2020: 19165-2 - Geographic information — Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products, https://www.iso.org/standard/73810.html

Ramapriyan, H. K., and P. J. T. Leonard. 2020. Data Product Development Guide (DPDG) for Data Producers version1. NASA Earth Science Data and Information System Standards Office, 21 October 2021. https://doi.org/10.5067/DOC/ESO/RFC-041VERSION1

Ramapriyan, H.K., Moses, J.F., & Smith, D. (2022a). NASA Earth Science Data Preservation Content Specification, Revision C. NASA Earth Science Data and Information System Standards Coordination Office, 3 May 2022. doi:10.5067/DOC/ESO/423-SPEC-001

Ramapriyan, H.K., Moses, J.F., & Smith, D. (2022b). Preservation Content Implementation Guidance, Version 1.0. NASA Earth Science Data and Information System Standards Coordination Office, 25 January 2022. doi:10.5067/DOC/ESO/RFC-042

USGCRP, 1998: Global Change Science Requirements for Long-Term Archiving, Report of the Workshop, October 28-30, 1998, National Center for Atmospheric Research, Boulder, CO, sponsored by NASA and NOAA, through the USGCRP Program Office.

Many Thanks!

Francis Lindsay and the ESDIS Team
francis.Lindsay-1@nasa.gov
earthdata.nasa.gov

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE