

Minutes of the 58th Meeting of the CEOS Working Group on Information Systems & Services (WGISS)

Chaired by USGS
Sioux Falls, South Dakota, USA

Table of Contents

Day 1: Wednesday 16th October, 2024	3
Session 1: Opening Session	3
1.1 - Welcome & Review of Agenda	3
1.2 - Data Preservation & Stewardship Interest Group (DSIG) Work Plan	3
1.3 - Data Discovery & Access Interest Group (DAIG) Work Plan	3
1.4 - Data Interoperability & Use Interest Group (DIIG) Work Plan	4
1.5 - Technology Exploration Interest Group (TEIG) Work Plan	4
1.7 - GEOSS Infrastructure Development Task Team	5
Session 2: Collections Management (DSIG)	7
2.1 - Session Introduction & Status of Collections Management Work	7
2.2 - Archiving Policy at ESA	7
2.3 - Archiving Policy at NOAA	9
2.4 - Archiving Policy at DLR	11
2.5 - Archiving Policy at JAXA	12
2.6 - Archiving Policy at USGS	14
2.7 - Collections Management Best Practice Review and Discussion (continued)	15
Session 3: Communications	15
3.1 - Discussion: Approval process for WGISS documents	15
1.6 - Select Action Review	16
3.2 - Discussion: How to effectively communicate WGISS Activities?	17
Session 4: Heritage Datasets Recovery (DSIG)	18
4.1 - Status of AVHRR data recovery work	18
4.2 - FDR4AVHRR	20
4.3 - Differences between AVHRR L1b/1c satellites products	21
4.4 - Australian AVHRR update	22
Session 5: Other Data Preservation and Stewardship Topics	23
5.1 - NOAA Collection Metadata Editing Tool (CoMET)	23

5.2 - WGISS DSIG Work Plan 2025-2027	25
5.3 - Software preservation White Paper - starting point for discussion	26
Day 2: Thursday 17 October, 2024	28
Session 6: STAC Topics	28
6.1- Introduction to STAC in WGISS	28
6.2 - INPE STAC Catalog integration in FedEO	29
6.3 - CEOS Service Discovery BP - STAC integration	30
6.4 - ISRO STAC overview	31
6.5 - STAC Summit overview and outcomes	32
Session 7: Federated Authentication and Authorization	34
7.1 - Introduction to Federated Authentication and Authorization	34
7.2 - Integrity Provenance and Trust - OGC Testbed-20 Report	37
7.4 - Identity Federations at DLR EOC	37
7.3 - Federated Integration of INPE Datasets	38
Session 8: Technology Exploration	39
8.1 - AI/ML White Paper	39
8.2 - Jupyter Notebooks Best Practices	41
8.3 - Knowledge graphs	41
Session 9: Challenges with Data in the Cloud	43
9.1 - Review of CGMS Best Practices	43
9.2 - Pain points identified by NASA	44
9.3 - Digital Earth experiences	45
Session 10: Closing Session	46
10.1 - WGISS-58 Action Review	46
10.2 - Closing Remarks	47
Appendix A: List of Participants	48
Appendix B: Actions	49

Day 1: Wednesday 16th October, 2024

Session 1: Opening Session

1.1 - Welcome & Review of Agenda

Tom Sohre (USGS, WGISS Chair) reported [[slides](#)]:

- Welcomed all to WGISS-58, and invited all to introduce themselves.
- Reviewed key points from the joint meeting with WGCV yesterday.
- Asked all to consider the following questions throughout the meeting:
 - Do we have the resources necessary to accomplish this work?
 - Is there additional work that we need to consider?

Discussion

- Steve Covington (USGS) recommended CEOS look at search engine optimization for the CEOS website, to ensure the outputs are visible online to the community.

1.2 - Data Preservation & Stewardship Interest Group (DSIG) Work Plan

Mirko Albani (ESA) reported:

- DSIG closed two actions from the 2021-2023 work plan [[slides](#)]:
 - DATA-22-04: Data Management and Stewardship Maturity Matrix
 - DATA-22-06: Archive Technologies White Paper
- For the 2023-2025 work plan, DSIG is working on two items:
 - White paper on EO Data collections management and governance
 - White paper on software preservation
- To get involved in DSIG work, contact Mirko or Iolanda Maggio.

1.3 - Data Discovery & Access Interest Group (DAIG) Work Plan

Damiano Guerrucci (ESA) reported [[slides](#)]:

- DAIG is working on DATA-22-05: Feasibility Study for Common Guidelines for STAC Implementation. The deadline for this will likely need to be extended.

- Other work is around Federated Authentication and Authorization, which will be discussed tomorrow, and the ongoing task of connected data assets and data discovery registration.

Discussion

- Tom Sohre (USGS, WGISS Chair) asked that when Federated Authentication is discussed tomorrow, the definition of the various terms and what is meant in this case is also discussed.

1.4 - Data Interoperability & Use Interest Group (DIIG) Work Plan

Nitant Dube (ISRO, WGISS Vice-Chair) reported [[slides](#)]:

- Last year, the Interoperability Framework was completed. The Interoperability Handbook version 2 is now under development, and will be completed by the end of 2025.
- In 2025, DIIG will also look into developing an Interoperability Maturity Matrix. Following this, in 2026, the team will develop some demonstration activities, including Earth Observation Plug & Play (EOPnP) modules, and further develop the CEOS Common Online Dictionary.

Discussion

- Steven Ramage (CEOS Executive Officer) noted there are similar activities happening within the World Economic Forum, as well as GEO's Earth Intelligence Readiness Matrix. The team should ensure to review this work to avoid duplication.

1.5 - Technology Exploration Interest Group (TEIG) Work Plan

Yousuke Ikehata (JAXA) reported [[slides](#)]:

- TEIG is working on two deliverables for 2024:
 - o DATA-23-01: AI/ML White Paper
 - o DATA-22-01: Jupyter Notebook Best Practice
- Also looking at topics related to Federation, including distributed data centres and APIs.
- TEIG is looking for a new co-lead to support Yousuke, as Richard Moreno (CNES) has recently changed roles and stepped down as co-lead.
- The scope should also be updated to reflect new technology topics.

1.7 - GEOSS Infrastructure Development Task Team

Paola De Salvo (GEO Secretariat) reported [[slides](#)]:

- The GEO Data and Knowledge Working Group has a number of subgroups:
 - o The In-situ data subgroup has the overall objective to support continued development and promotion of a strategy for GEO that addresses the barriers to sharing and re-use of in-situ data. The GEO In Situ Data Strategy is in development and open for comment.
 - o The Data Sharing and Data Management Principles subgroup is working on a number of activities including development of Data Management Principles, Data Sharing Principles, implementation Guidelines and the Self-assessment tool Dialogue series.
 - o The Law & Policy subgroup has the objective to engage with GEO Work Plan Activities to better understand and leverage open data licences.
- The GEO Data and Knowledge Working Group will submit to be a convener under the new GEO Post-2025 Work Programme.
- The Post-2025 Work Programme is centred around six thematic Focus Areas:
 - o Agriculture and food security
 - o Water and land sustainability
 - o Ecosystems, biodiversity and carbon management
 - o Weather and disaster resilience
 - o Climate, energy and urbanisation
 - o One health
- The GEO Infrastructure Development Task Team is working to develop a common user interface to access both the GEO Knowledge Hub and GEOSS - building on existing work. The linkage with regional and national GEOs will also be strengthened.
- The GEO infrastructure should be a place where a user can find resources organised by categories to support specific challenges and applications.
- The [GEO Knowledge Hub](#) is a central library where resources are organised by the relevant community. The resources are organised into 'Knowledge Packages', which can include in-situ data, space-based data, documents, user stories, code / software, etc. It currently holds 154 packages and 816 resources.

- Webinars have also been held to demonstrate the knowledge hub, and interact with the user communities. A centralised place for resources is extremely valuable for the users.
- GEO is planning to continue this engagement, especially with the youth community. This was highlighted at the Open Data Open Knowledge (ODOK) workshop held recently in China.
- The Space-Based data session at ODOK 2024 produced some actions, including:
 - o Advocate for greater adoption of open standards and open data cubes in the GEO community.
 - o Build on existing [CEOS] Analysis Ready Data specifications
 - o Adopt developing technologies (Cloud and Quantum Computing, Docker, Kubernetes, Deep Learning).
 - o Enhance capabilities for joint calibration and validation of remote sensing data & products. See QA4EO and CEOS WG CalVal.
 - o Strengthen regional collaboration and capacity building around space-based data in GEO.
 - o Support open data policies and efforts and technologies based on CARE, FAIR and TRUST

Discussion

- Tom Sohre (USGS, WGISS Chair) noted that WGISS could look at supporting some of those actions.
- Nitant Dube (ISRO, WGISS Vice-Chair) asked about what interoperability resources are available.
- Paola De Salvo (GEO Secretariat) noted that the team is in the process of reassessing the quality of metadata provided by data providers, to improve the discoverability, accessibility and reusability. The Working Group is eager to strengthen the connection with work done by CEOS, and ensure to not duplicate.
- Steven Ramage (CEOS Executive Officer) suggested CEOS pick a few items from the ODOK Workshop actions which align with the CEOS Work Plan.
- Paola noted that the GEO Executive Committee has asked GIDTT to present in April a list of resources required to develop the suggested infrastructure. The goal is to build something more attractive for the end user to find the datasets and information they need. A few case studies will be developed, to prove it is a good way forward.

WGISS-58-01	Interest Group leads to review the GEO Open Data Open Knowledge (ODOK) actions to identify opportunities for collaboration.	Due: Q4 2024
--------------------	---	-------------------------

Session 2: Collections Management (DSIG)

2.1 - Session Introduction & Status of Collections Management Work

Mirko Albani (ESA) reported [[slides](#)]:

- The goal is to manage and mitigate challenges when managing collections, including integrity, authenticity, replica management, reproducibility, citation and archiving.
- The team is working to produce a White Paper on EO Data Collections Management and Governance.
- The document is in good shape, and is open for comments and review.
- There will be a version 1.0 produced and published, but will be regularly reviewed and updated when new developments occur.
- The session today focuses on archiving policies for collections, and aims to gather feedback from WGISS members on their procedures.
- The development of the document will occur through Google Docs.

WGISS-58-02	Iolanda Maggio to circulate the latest version of the Collections Management document to the writing team, WGISS Exec & WGISS-58 participants for final review.	Due: 31st October
--------------------	---	------------------------------

2.2 - Archiving Policy at ESA

Mirko Albani (ESA) reported [[slides](#)]:

- The 3-2-1 rule (three copies, two formats, one off-site) is used at ESA to ensure the integrity of the archives.
- There is an increasing volume of data, and hence it is not sustainable to archive everything. The challenge is to understand the correct and minimal way to store archives.
- The ESA EO Space and Ground Data Archiving Policy covers all EO missions (excluding Sentinels) archived at ESA. The packages are composed of:
 - o Space-borne and ground segments datasets (SGD)

- Software, analysis tools and services data
 - Technical and Scientific Information
 - Metadata
- Data is grouped based by their processing levels into four categories: raw data, native format data, reformatted data, repackaged data. Raw data is always kept in case there are issues in the processing.
 - ESA's archiving policy is as follows:

Payload Raw Data

- Payload raw data shall be archived in **2 (two) copies in 2 (two) physical locations** distant at least 200 Km, using 2 (two) different archiving technologies.

Native Format Data: Payload L0, spacecraft telemetry, auxiliary and ancillary

- **Current Baseline** of Master datasets shall be archived in **3 (three) copies in 3 (three) physical locations** distant at least 200 Km among them, and using 2 (two) different archiving technologies. If a corresponding repackaged dataset is available, the current baseline of Master datasets shall be archived in 2 (two) copies in 2 (two) physical locations distant at least 200 Km, using 2 (two) different archiving technologies.
- **Previous Baseline** of Master datasets, if existing, shall be archived in **2 (two) copies** in 1 (one) or more physical locations, using 2 (two) different archiving technologies.

Reformatted and/or Repackaged Data: Payload L0, spacecraft telemetry, auxiliary and ancillary

- **Current Baseline** of bulk reformatted datasets, if existing, shall be archived in **2 (two) copies** in 1 (one) or more physical locations, using 2 (two) different archiving technologies.

Native Format Payload Higher-level data (e.g. Level-1/Level-2 bulk reprocessed)

- **Current Baseline** of bulk processed/reprocessed and quality approved datasets shall be archived in **3 (three) copies** in 3 (three) different physical locations, distant at least 200 Km among them using at least 2 (two) different archiving technologies, **if reprocessing from previous level (e.g. Level-0) is not achievable anymore** for technical or cost reasons. If a **repackaged** corresponding dataset is available, the Current Baseline of bulk processed/reprocessed datasets shall be archived in 2 (two) copies in 2 (two) physical locations distant at least 200 Km, using 2 (two) different archiving technologies.
- **Current Baseline** of bulk processed/reprocessed and quality approved datasets shall be archived in **2 (two) copies** in 1 (one) or more physical locations using 2 (two) different archiving technologies **if reprocessing** from previous level (e.g. Level-0) is demonstrated to be still **achievable** both at technical and cost level
- **Previous Baseline** of bulk processed/reprocessed and quality approved datasets shall be archived in **2 (two) copies** in 1 (one) or more physical locations using 2 (two) different archiving technologies.
- Higher-level products generated **on-demand** shall **not be permanently archived** (e.g. on-the-fly production).

Reformatted and/or Repackaged Payload Higher-level Data

- **Current Baseline** of bulk reformatted and/or repackaged and quality controlled datasets, if existing, shall be archived in **2 (two) copies** in 1 (one) or more physical locations, using 2 (two) different archiving technologies.

- Software, analysis tools and services data including relevant documentation related to EO Space and Ground Segments Datasets shall be archived in at least **3 (three) copies** in 2 (two) different locations distant at least 200 Km using 2 (two) different technologies.
Archived SGD related software and data analytics tools shall contain all elements needed to properly run them. As an example, instrument processing facilities shall contain processors, orchestrators and any required software and configuration file (e.g., Task Tables) to properly set up the processing environment. Ready to run preconfigured processing environment (e.g., based on containerized SW) is to be preferred.
- Technical and Scientific Information related to EO Space and Ground Segments Datasets shall be archived in at least **3 (three) copies** in 2 (two) different locations distant at least 200 Km using 2 (two) different technologies.
- Metadata describing the context, content and structure of the digital objects composing Space and Ground Datasets Packages shall be archived in at least **3 (three) copies** in 2 (two) different locations distant at least 200 Km using 2 (two) different technologies.

- This can result in up to eight copies - in particular for the heritage missions. ESA is looking at preserving the data processing algorithms so reproducibility is possible.

Discussion

- Tom Sohre (USGS, WGISS Chair) suggested more guidance about using the cloud for archives may be useful. If the 'off-site' copy is on the cloud, it should be a different cloud to the primary copy - specifically a geographically different cloud.
- Robert Fletcher (UKSA) noted egress costs to restore data from the cloud could be through the roof.
- Mirko added that ESA has copies for dissemination which are not included in the three archive copies. ESA does not disseminate from the archive.
- Sai Kalpana (ISRO) recognised that if the backups aren't organised in a similar manner then restoring can be more difficult. How you get back the data in case of an issue with the dissemination copy is important.
- Sai asked how frequently are the backups generated?
- Mirko noted this depends if reprocessing is needed. A master copy is taken from the archive when reprocessing is done, and then the reprocessed dataset is returned if applicable. If a new baseline is created, the previous baseline is kept as well.
- Doug Newman (NASA) suggested that the metadata copies could be reduced if they can be reproduced from the original data.
- NASA is looking at who else has copies of their data and who is archiving it, especially with interagency activities. However, this only mitigates risk if there is an agreement in place between the agencies.

2.3 - Archiving Policy at NOAA

Ken Casey (NOAA) reported [\[slides\]](#):

- With their recent experiences due to Hurricane Helene, it has highlighted the fact that agencies should be aware of how long data producers can hold the data if the system goes down. NOAA has agreements with the producers to hold the data for two weeks, but as the recent outage was approaching the two week window, they had to confirm exactly how long the data could be kept. NOAA will review this in the coming months.
- NOAA uses 'granule' to refer to the smallest discoverable unit, 'dataset' as an aggregation of granules based on similarity, and 'collection' as the highest level of aggregation that is citable.

- Federal Records Management law requires Federal Agencies to document policies and transactions and adhere to National Archives and Records Administration (NARA) guidelines.
- NOAA also has a records retention schedule which in collaboration with NCEI, has specific guidance for some environmental data. NCEI develops Archive Recommendation Packages for NOAA programs that define what to archive and how long to retain
- Original observations are preserved permanently, while oceanographic and geographical data are preserved for 75 years.
- After ten years, products and product versions, related documentation and metadata are reassessed for community needs prior to disposition.
- NCEI stores two copies of all digital data, and one copy is geographically distanced from the data centre. The cloud archive is in one region, but with multiple copies. However, they are all within 100km of each other on separate power and cooling etc. This wouldn't be resilient to major geological disasters, such as a tsunami.
- NOAA's archive policy is focused on open science and scientific integrity and reproducibility.

Discussion

- Peter Cornillon (University of Rhode Island) asked about the availability of copies on site.
- Ken noted NOAA currently has multiple archive systems. There are three on-premise archive systems, with each having data at the main location (Boulder or Ashfield) and in a remote location. All of the CLASS data are now in the cloud, but not integrated with the NESDIS common cloud framework. NOAA is currently working to migrate all data to a new cloud archive system.
- The current intent is to have everything archived on the cloud, but discussions around cost and risk are evolving.
- NOAA archives all levels of data, level 0 and levels 1-4.
- Makoto Natsuisaka (JAXA) recognised there is a need to archive software and programs as well, but hard to store long term as technology changes.
- Ken noted that within NOAA, the archiving and public sharing of software is not done consistently across the programs. NOAA is in the final stages of publishing a new software sharing policy which uses the NOAA central library to ensure all software is

shared appropriately with the world. Over time, NOAA should get to a more consistent approach. There is also the need to put sufficient protections in place to make sure a commercial entity doesn't try to copyright the software.

- Robert Fletcher (UKSA) asked about the level of data volumes stored on the cloud, and the AWS technologies used.
- Ken noted that NOAA has a couple PB of data in the cloud, but in the formal archive there is over 30 PB of data. The current plan is to keep a hot copy on S3 low latency, with a 20-30 day hold at this level. At the same time, this data is copied to Amazon deep storage. User action patterns drive and dictate what is kept on low latency storage. NOAA has a public-private partnership with Amazon to use the open data dissemination system, where NOAA data is stored on three cloud services for access, on S3 low latency at no cost to NOAA.

2.4 - Archiving Policy at DLR

Katrin Molch (DLR) reported [\[slides\]](#):

- DLR data management objectives are to preserve, curate and ensure long-term accessibility and usability for data of national missions and of DLR interest. They facilitate EO data access and use for DLR scientists, including via online storage with processing capabilities.
- Data can be disseminated from the archive, but this is done less and less with cloud services. Bulk data can be disseminated as well.
- The archiving policy at DLR is complex. For national EO missions, all levels of data are currently stored, with no time limit. Two copies are in each of Oberpfaffenhofen and Neustrelitz DLR centres.
- For international EO missional data of DLR scientific interest, two copies are held at one site, in different buildings and using different technologies.
- A number of datasets are also maintained on contract on behalf of other data providers. The policy differs, but for example, for GOME-2, Level-2 reprocessing is done about every 5 years, and the previous version maintained until the reprocessed version is validated. A 3 version history is maintained in the Technical Note.
- For significant in-house spatial information products, two copies are held at one site in different buildings.

- Copernicus Sentinel user level data of DLR scientific or commercial interest is also preserved at DLR. Most product types in complete time stacks are stored. Only one copy is stored in the long term archive.
- For Sentinel-2 Collection-1, the plan is to store Level-1C data in the archive, and stepwise retire & remove previous L1C data from the long term archive (~11 PB). Level 2 data would only be available in the analytics platform.
- For Sentinel-1 and Sentinel-2, the data volume is getting very large, and DLR would welcome cross-agency coordination to help manage the archiving of these datasets.
- For taxonomy, DLR's 'Collections' contain 'Products' with 'Components' which are available in 'files'.
- DLR is approximately 80% compliant with the CEOS Data Preservation Guidelines.
- DLR does not have yet any plans regarding preservation of technical documentation and/or relevant software, however it is getting urgent for TerraSAR-X and TanDEM-X missions.

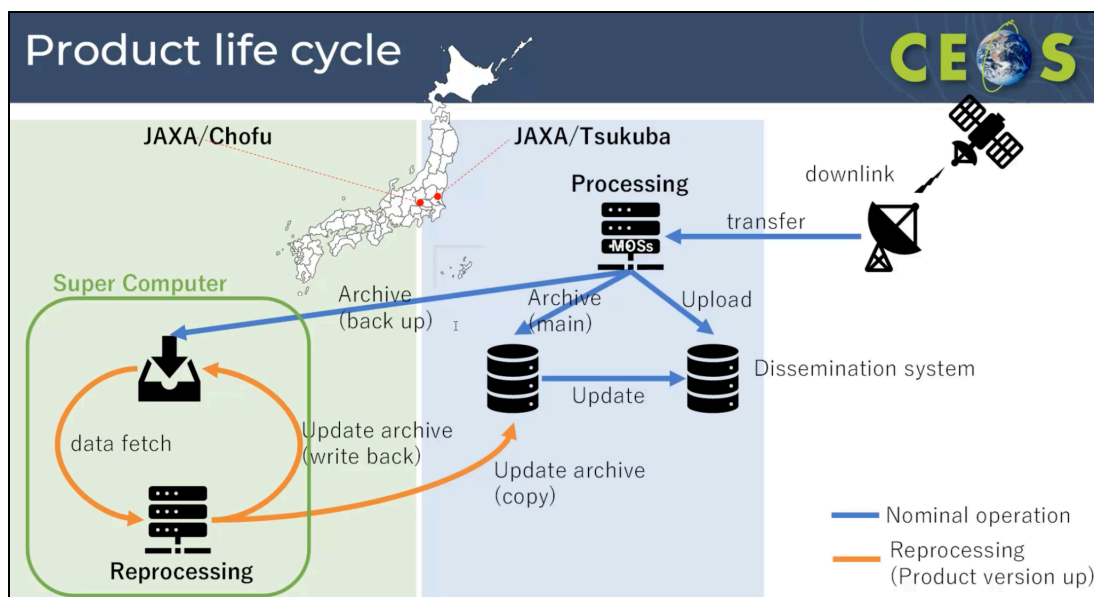
Discussion

- Sai Kalpana (ISRO) asked about the purpose of the different technologies for the two copies of the international EO data.
- Katrin noted that one copy is held on tape, and another one a more efficient technology on disk. There is also built-in redundancy in the disk system as well.
- Peter Cornillon (University of Rhode Island) asked whether two different versions of a level 2 dataset are archived in the same way? For example, two different NDVI algorithms.
- Ken Casey (NOAA) noted that NOAA has real-time satellite datastreams for generating L2 products which are archived. Later on, the CDR team reprocesses these to make more accurate and consistent datasets which are also archived. However, when the same dataset goes through multiple reprocessings, NOAA may only keep the most recent few - there is no strict policy.
- Steve Covington (USGS) noted that the Level 1 and 2 near-real time products generated by USGS are replaced when the higher fidelity product is ready.

2.5 - Archiving Policy at JAXA

Yousuke Ikehata (JAXA) reported [\[slides\]](#):

- JAXA defines 'Master Data' as those which can't be obtained nor reprocessed if lost. These data must be double archived, alongside the processing code and related documents.
- 'Products' are defined as those processed from master data, and are mostly archived in mission operation systems (MOSSs) and data dissemination systems like G-Portal, AUIG-2 (ALOS and ALOS-2), Bousai-IF (disaster-related products), etc. to provide those to users.
- JAXA doesn't use the cloud for archiving - availability covers only the system, and hence double systems are used.
- JAXA stores and manages the master data without a time limit. The data is mirrored in 2 data centres to avoid disaster.
- Processing systems aren't typically preserved, the preservation of documents describing processing software is decided by each project.
- Without the master data, the data is only stored for a fixed term - however long users may want to access the data. This is almost always an indefinite term. The data is only stored at one data centre.



- Tsukuba data centre carries out real time processing and acts as the archiver for hot storage. Tsukuba is also in charge of the dissemination system. Chofu does re-processing and archives for cold storage.
- Shared storage systems at Tsukuba are replaced about every 5 years.

Discussion

- Steve Covington (USGS) asked whether agencies should be moving away from the number of copies needed, and instead focus on the reliability of the copies to ensure survivability of the master archive.
- Tom Sohre (USGS, WGISS Chair) noted that USGS uses an availability metric to understand reliability.
- Doug Newman (NASA) recognised that NASA used a similar metric for durability, but it doesn't account for someone deleting the data.

2.6 - Archiving Policy at USGS


Ryan Longhenry (USGS) reported [\[slides\]](#):

- Data is preserved using long-term archive capabilities on multiple tape copies of differing technologies, that are geographically distanced.
- The National Satellite Land Remote Sensing Data Archive (NSLRSDA) was created to identify, collect, preserve, and make available land-based, remotely sensed imagery. The collection policy is based on the concept of maintaining imagery collections that serve environmental, agricultural, mapping, energy, economics, and global change research communities in addition to the general public.

Data Archiving Procedures


- Data is preserved using 3 copies: nearline, offline, and offsite
 - The most "native" level is deemed as the archival version in most cases
- Nearline data stored in a Spectralogic Tfinity robotic tape library system
- Offsite copies are stored at the National Archive Record Administration (NARA) Kansas City facility
- Multiple media types are utilized to ensure long term readability
 - Currently using LTO-9 and IBM TS1160 tapes
- Media are periodically updated as tapes technology advances.
- Records management schedules are worked with NARA

nearline




at EROS

offline



offsite



at NARA

Discussion

- Sai Kalpana (ISRO) asked how the storage process differs for decommissioned missions.
- Ryan noted that USGS was given all of the raw data from EO-1, alongside the processing system. Since USGS weren't planning to operationalise, as part of the deactivation of the mission, the highest processed Level 1 data was also stored. For active missions, USGS continues to migrate to new technologies.

- Commercial providers which decommission their mission and don't want to store the data have to legally offer it to the US government first. The data can be sometimes in a proprietary format, so USGS stores level 0 data.

2.7 - Collections Management Best Practice Review and Discussion (continued)

Mirko Albani (ESA) reported [[slides](#)]:

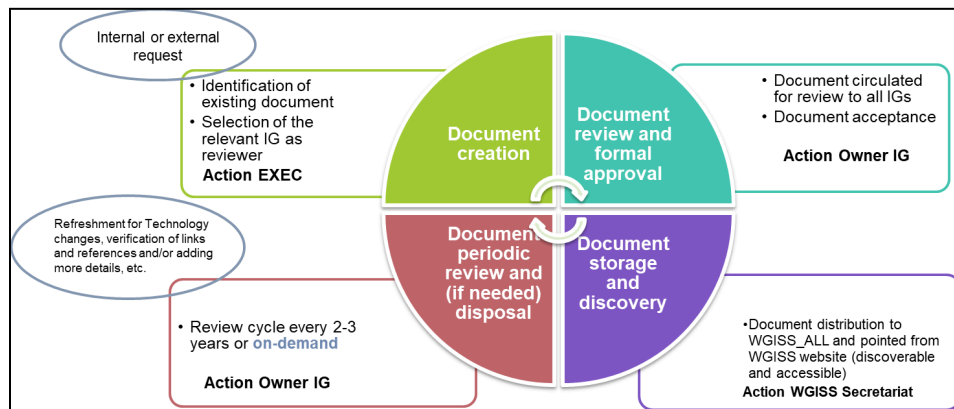
- The document is in good shape, and available [here](#).
- This will be shared with all participants for review.
- The archiving presentations from today form a small part of the collection management best practices.
- Once the document has been finalised, DSIG will also review associated documents previously published.

Session 3: Communications

3.1 - Discussion: Approval process for WGISS documents

Iolanda Maggio (ESA/Starion) reported [[slides](#)]:

- At WGISS-57, it was decided that WGISS web pages should be reviewed every two years once a new chair commences their term.
- The following document review cycle was also agreed:



- There is some ambiguity around the approval step, who has to review it and formally approve it.

Discussion

- Iolanda confirmed the steps for the Collection Management white paper would be to share with Exec for a two week review period, before collecting any comments and finalising the document.
- Steven Ramage (CEOS Executive Officer) suggested that broad things such as the Interoperability Handbook should be brought to CEOS Principals at SIT or Plenary for endorsement. In particular, any document which should be distributed widely under the CEOS banner.
- Tom Sohre (USGS, WGISS Chair) recognised the difference would be any resourcing request, in particular for groups other than WGISS. Any high level strategic documents should also go to plenary. Technical documents internal to WGISS should not require external endorsement.
- Steven noted that the CEOS Work Plan is approved annually by Principals, and includes details on all the work ongoing within Working Groups.
- Nitant Dube (ISRO, WGISS Vice-Chair) requested standardised nomenclature for the types of documents which should go to plenary and or not.
- Libby Rose (WGISS Secretariat) suggested this approval process should be brought to CEOS Secretariat for discussion, to ensure CEOS management agree with the agreed process.

WGISS-58-03	DSIG to update the document governance process to define that approval by CEOS Plenary is only necessary for high-level strategic documents, especially where it is a collaborative effort across CEOS, or otherwise where deemed appropriate when submitted to work plan. This should then be brought to CEOS SEC by the WGISS Chair, to ensure it aligns with broader WGISS processes.	Due: 15 November 2024
--------------------	--	--------------------------------------

1.6 - Select Action Review

Libby Rose (WGISS Secretariat) reported [\[slides\]](#):

- There are currently 5 open actions, and 16 in progress in the [WGISS Action Tracker](#).
- Action WGISS-57-08 was recorded to update the Terms of Reference.
- The changes were presented to WGISS Exec last month, with the main change in the stakeholder section. The language aligns with the external engagement language proposed by the SEO for the CEOS website.
- This will be presented to CEOS Plenary next week for formal endorsement.

Discussion

- No objections to the changes to the Terms of Reference were raised.
- Tom Sohre (USGS, WGISS Vice-Chair) suggested that the Terms of Reference be reviewed at a regular cadence.

3.2 - Discussion: How to effectively communicate WGISS Activities?

Libby Rose (SEO Comms Team) reported [[slides](#)]:

- CEOS has a number of channels which can be used by WGISS to communicate their activities.
- Exhibition booths at external conferences are a great way to interact one-on-one with the community. WGISS can contribute materials for these, in the form of videos, slides, or hardcopy materials.
- The key question to understanding how to effectively communicate WGISS Activities is around the goals of WGISS, and the target audience for each piece of work. Where the target audience.
- The website is largely structured around the working groups. There is also a dedicated publications page, as well as the 'our work' page which has links to tools such as IDN and FedEO.

Discussion

- Steve Covington (USGS) suggested a webinar to introduce CEOS activities to the broader community.
- Tom Sohre (USGS, WGISS Chair) proposed that the goal of WGISS communications should be to better tell the story of what WGISS is doing for the community.
- Damiano Guerrucci (ESA) noted that WGISS should try to first communicate to the broader community, then show that to management, including the communication piece.
- Libby suggested that, for example, for the Collections Management White Paper, a high level summary be written as a blog post on ceos.org/news, which can then be shared via social media. But the detailed strategy would depend on the target audience.
- Tim Stryker (USGS) noted the Earth Science Information Partnership within the US. Can WGISS coordinate with them, and other similar organisations to spread the word?
- Steve suggested CEOS could host a session at a future IGARSS or similar conference. However, there may be issues with funding for this.

- Tom noted that many WGISS members attend these types of conferences on behalf of their agencies, and could spread the WGISS messaging as well.
- There is also the need for an up-to-date set of CEOS slides, which can be pulled out when needed. The Executive Officer does maintain a set of slides, however it can be hard to make sure the slides are always up to date, and also at a correct level for the relevant audience.
- Steven Ramage (CEOS Executive Officer) recommended that when communicating activities, WGISS should consider what the audience should 'do' - whether that is review documents, act on recommendations, etc. This piece is currently missing.
- Tim suggested the broad action should be for data providers to make their data discoverable, accessible and usable.
- Steven also suggested campaigns of communication activities, to ensure repeated engagement of the audience with the topic.

WGISS-58-04	WGISS Secretariat to discuss with the SEO about increasing the impact of information posted on the CEOS website.	Due: WGISS-59
WGISS-58-05	In developing the 2025-2027 CEOS Work Plan, Interest Group leads should identify the target audience(s) for their deliverable to inform the communication approach.	Due: Q1 2024

Session 4: Heritage Datasets Recovery (DSIG)

4.1 - Status of AVHRR data recovery work

Mirko Albani (ESA) reported [[slides](#)]:

- AVHRR Local Area Coverage (LAC) data is available in a global archive from EUMETSAT from March 2008 onwards. However, many national / regional data archives of LAC data around the world covering a longer period with high value for the retrieval of ECVs. Some of them accessible to users, others not due to unknown accessibility, responsibility, data format and structure.
- The objective of the project is to build a worldwide coverage AVHRR LAC data series from 1978 onwards.
- Two different global Land 1 km AVHRR datasets covering the period 1992-1999 are available at ESA, one with data acquired by ESA network stations (Terranova, Nairobi, Manila, etc..), and the other with data acquired by USGS network stations, however

these were in stitched format (.arch files). With some assistance from CSIRO, a tool has been found to read this data.

- The European dataset includes data from University of Bern, Dundee Station and ESA holdings. The 260,000 products were harmonised and consolidated through a dedicated ESA project (Heritage Space Programme). Processing to Level-1C has been completed, and data will be opened to users in Oct 2024.
- A new activity was started in Q1 2024 to reprocess the AVHRR series LAC products (NOAA & MetOp missions) to produce a Fundamental Climate Data Record for over 40 years.
- ESA contacted the IGIK (Polish) point of contact in 2023 and 2024 but haven't received a response.
- USGS and NRCAN will be invited to make a presentation at WGISS-59 on the status of their AVHRR archive.
- Data transfer to ESA from the Argentinian archive is ongoing. Downloading was stopped in April 2024 to upload new data in the GiDyC servers, restarted in October.
- Data transfer to ESA from the Brazilian archive has been completed.
- All exabyte tapes and hardware in the Hawaiian archive were shipped to ESA/ESRIN, and the transcription chain is being assembled. The ESRIN laboratory is manufacturing a special device to roll/unroll the tapes at very slow speed and clean the tape surface from moisture.
- The data transfer to ESA from SANSA has been completed.
- ESA has successfully transcribed the DLTs from Kenya.
- ESA sent a request for accessing the data from China. CMA will be contacted and invited to make a presentation at WGISS-59.
- The Mongolia institute will be contacted to get information on data volume and status of accessibility
- A list of Meteorology offices and other organisations around the world who might have AVHRR LAC data is being compiled at ESA.
- ESA will perform a worldwide AVHRR LAC data gap analysis extending the one done for Europe, and contact additional organisations who might have AVHRR LAC data to investigate possibility to fill identified gaps
- The next steps include:

- Continue ongoing transcription activities, coordination and projects and contact additional organisations with potentially unique AVHRR LAC data
- Provide AVHRR points of contact and stakeholders with the AVHRR Dataset Consolidation Procedure and Processing Format used in the European consolidation project run at ESA. Collect feedback and discuss possibility to have a common/harmonised format (Q1/Q2 2025)
- Check status of Discoverability of AVHRR datasets worldwide and pursue common entry point through WGISS CDA infrastructure (Q1 2025)

Discussion

- Sai Kalpana (ISRO) noted that ISRO has the level 0 data on tapes, but the products are not available for dissemination yet. If ESA can deliver the software, ISRO can process and disseminate.

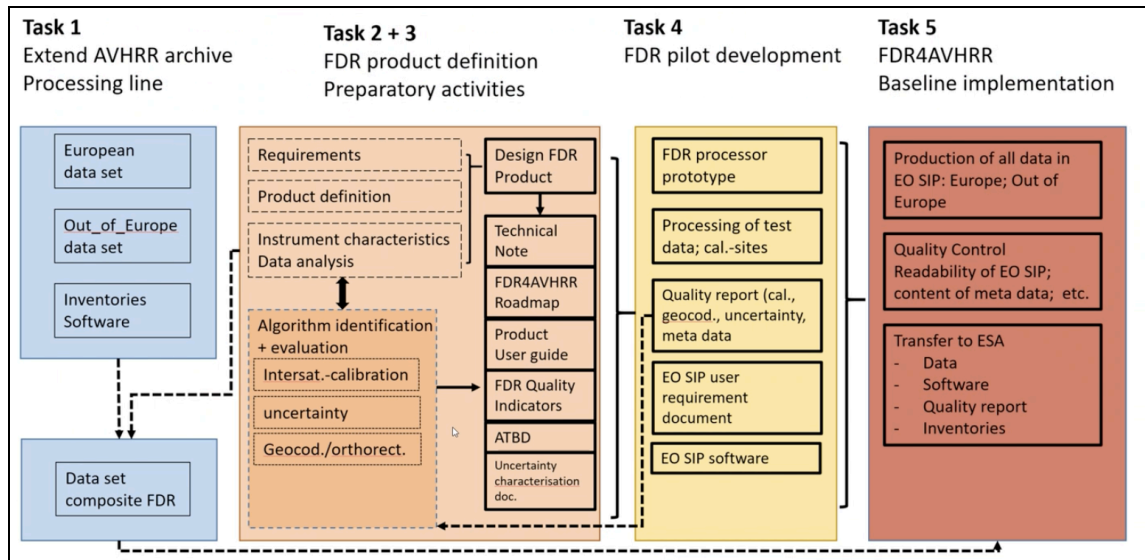
WGISS-58-06	Mirko Albani to contact ISRO about the processing of their AVHRR data and contribution to the AVHRR recovery activities under DSIG lead.	Due: 15 November 2024
WGISS-58-07	Mirko Albani to connect with SST-VC regarding presenting the DSIG AVHRR work at GHRSSST, including for potential future activities on the heritage data recovery topic.	Due: WGISS-59

- Peter Cornilion (University of Rhode Island) noted that CLASS has a lot of LAC data, including over Russia. Will likely have a lot more over north America.

4.2 - FDR4AVHRR

Stefan Wunderle (University of Bern) reported [\[slides\]](#):

- Data from the University of Bern, ESA stations and Dundee was reprocessed to Level-1B (standard NOAA format).
- The FDR4AVHRR curation project is aiming to generate a fundamental data record (FDR) based on AVHRR data and to make all data accessible via ESA dissemination service. Global data will be reprocessed to Level-1C, allowing for higher level products to be developed.



- The overlap of the Dundee and UniBern archives allows for redundancy. The team will develop a procedure to support users with additional data, checking for redundancy and quality before a decision is made about integration into the dataset.
- Data collected from Argentina, South Africa and Kenya will be integrated into the FDR4AVHRR project.
- The validation strategy involves ensuring quality control through consistency, completeness and integrity of the FDR product, as well as validation of calibration and geocoding.
- At the end of the project (in Nov 2025), approximately 500,000 1-km AVHRR Level 1C data products will be transferred to ESA for later accessibility via ESA dissemination service.

Discussion

- Peter Cornillion (University of Rhode Island) asked about the procedure for geocoding.
- Stefan noted this is based on a reference mosaic, the blue marble 2010 mosaic, with a feature matching approach. The procedure is implemented at UniBern, including the elevation model. However this can't be done over open ocean. The PyGAC software also does the clock corrections first.
- Stefan is in contact with Edward King (CSIRO) who is helping ESA understand the stitched formats.

4.3 - Differences between AVHRR L1b/1c satellites products

Mirko Albani (ESA) reported [\[slides\]](#):

- During the session on AVHRR recovery project held during WGISS-57 it was requested to clarify the difference between the ESA L1b/1c, NOAA L1b/1c, EUMETSAT L1b/L1c, and CSIRO L1b/L1c data.

	ESA	NOAA	EUMETSAT	CSIRO
GAC L1b	N.A.	NOAA format	EPS format	TBD
GAC L1c	N.A.	N.A.	netCDF format	TBD
LAC L1b	NOAA format	NOAA format	N.A.	TBD
LAC L1c	netCDF format	N.A.	N.A.	TBD

- The next step is to extend the analysis and investigate potential approaches for GAC/LAV data product alignment.

WGISS-58-08	Matt Paget to share information with DSIG about the formats used by CSIRO for AVHRR data.	Due: 15 November 2024
--------------------	---	--------------------------------------

4.4 - Australian AVHRR update

Matt Paget (CSIRO) reported [\[slides\]](#):

- The Australian record currently goes back to 1992 April, and CSIRO is working on recovering data back to 1983.
- Public access to the AVHRR dataset is via the EASI Hub (using ODC). This data could be made available through the CEOS Analytics Lab if desirable.
- The data produced is a Level-1c netCDF product, and are looking at doing atmospheric corrections to produce BRDF and Surface Reflectance products.
- There have been encouraging preliminary results for comparison against MOD09 and MCD43A4.
- The next steps include evaluating the RTTOV approach with other methods (e.g., 6SV/MODTRAN6, CLAVR cloud masking, CAMS aerosol dataset), quantifying uncertainty in the retrieved SR and account for stratospheric aerosols (Pinatubo) .
- CSIRO will demonstrate time series analyses with notebooks, and add metadata and submit for assessment against the CEOS-ARD specifications.

Discussion

- Matt is not sure about the differences between the CSIRO dataset and the one generated over Europe. ESA and CSIRO should clarify this and present at the next WGISS meeting. At the moment they are two parallel activities.
- The format for CSIRO's Level 1B LAC product is netCDF.
- Stefan Wunderle (University of Bern) asked if it would be possible to recalibrate the stitched data. Can we use the stitched passes to improve the calibration?
- Matt noted this should be possible, as CSIRO has enough of the raw scenes and information to be reprocessed as we wish.

WGISS-58-09	Matt Paget & Stefan Wunderle discuss if/how the two AVHRR projects can converge, including on the re-calibration of data.	Due: WGISS-59
--------------------	---	--------------------------

Session 5: Other Data Preservation and Stewardship Topics

5.1 - NOAA Collection Metadata Editing Tool (CoMET)

Sarah Menassian (NOAA) reported [\[slides\]](#):

- NOAA's CoMET application enables users to perform a variety of tasks to create, manage, and assess collection metadata.
- CoMET is part of the overall Collection Manager Tools. It supports a number of functions, including creating / importing collection metadata records, validating and resolving XML metadata, and completing and evaluating a Data Stewardship Maturity Questionnaire (DSMQ) and receiving section-level scores.
- The metadata editor supports ISO 19115-2 standard.
- The data stewardship maturity matrix is a unified framework for measuring Stewardship Practices applied to individual digital earth sciences data products
- With DSMQ scores in the metadata, this can assist in machine automated selection of data fit for purpose.
- In CoMET, for collections with a DSMQ, the ratings are inserted into the collection metadata record. The overall score is displayed in NOAA's official data catalog, OneStop.
- The development team has been working to allow access to CoMET from outside NOAA. Once deployed, users who do not have a noaa.gov email address will be able to access

the application via login.gov. Upon initial login, CoMET will automatically create user's 'Personal Repository' in which records can be created, including DSMQs

- CoMET is now deployed to the NESDIS Common Cloud Framework (NCCF) development environment. The application will be fully switched to the cloud eventually, likely in more than 12 months time.

Discussion

- Peter Cornillon (University of Rhode Island) asked if the tool comes with best guesses for a specific dataset?
- Sarah confirmed that this isn't yet possible, however there are some templates. There are technologies available which could do this, but the team isn't working on this at the moment.
- Michael Morahan (NASA/KBR) asked when this will be available to non-NOAA users.
- Sarah noted a date is not yet clear, as the process for approval has been a bit lengthier than expected. NOAA encourages anyone to reach out who may be interested in the tool once released.
- Iolanda Maggio (ESA/Starion) noted the WGISS Data Management and Stewardship Maturity Matrix (DMSMM) was built off the CoMET one, but with different components.
- Mirko Albani (ESA) noted this isn't necessarily being adopted by CEOS, however CEOS might want to develop something similar in future. If we do this, it should be fully aligned with IDN / FedEO.
- Matt Paget (CSIRO) recognised it would be good to build on the CEOS-ARD product family specifications, and STAC metadata items, to develop a common online easy to use metadata building tool. Getting close to a dynamic buildable interoperable system.

WGISS-58-10	Anyone interested in using CoMET once it becomes available to non-NOAA users should send Sarah Menassian (NOAA) an email.	Due: WGISS-59
WGISS-58-11	DAIG to consider whether a CEOS tool should be created to develop a collection metadata file (ISO/STAC), similar to CoMET with maturity matrix support. Consider how this will connect with FedEO & IDN.	Due: WGISS-59
WGISS-58-12	NOAA CoMET team to discuss whether the tool will be open source, and consider how CEOS could build on this work to make a version applicable for CEOS.	Due: WGISS-60

5.2 - WGISS DSIG Work Plan 2025-2027

Mirko Albani (ESA) reported [[slides](#)]:

- DSIG enables the sharing of agencies’ investigations, developments, experiences, and lessons learned relating to EO data stewardship.
- The group would like to focus on five main topics over the upcoming period.
 - o Heritage data recovery: identify historical/heritage datasets currently not accessible to users and trigger potential joint recovery actions. In a follow up to the AVHRR work currently being done. WGClimate will be consulted for ideas on which datasets would be a priority.
 - o Archive holdings and technology (including interoperability between archives): share information on new technologies and approaches and draft best practices to facilitate archives upgrade and evolution. Understanding who else holds copies of an agency’s data would be good to investigate.
 - o Data associated information and software preservation: share information on approaches and technologies used for technical content and software preservation.
 - o Collections management and governance: share information on collections management and governance approach and define Best Practices/Guidelines
 - o Interaction with standardisation bodies: share information on international standards relevant for WGISS activities and contribute to standardisation activities. Work CEOS can complement the work done in standards bodies. Have lost the connection with OGC as of late.

Discussion

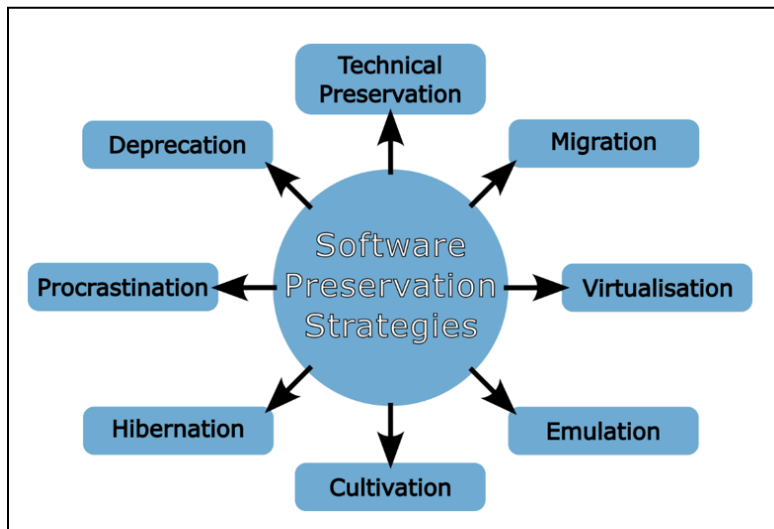
- Peter Cornillion (University of Rhode Island) suggested presenting the AVHRR work to GHRSSST as well.
- Katie Baynes (NASA) asked if the software preservation topic would be generalised for other domains, or just EO?
- Mirko noted the team will focus on just EO, focusing on what is needed to make sure the software needed to reprocess the raw data is preserved.

WGISS-58-13	WGISS Secretariat to come up with some language around some related resources for WGISS, and share with Exec to suggest which links should be included.	Due: WGISS-59
--------------------	---	--------------------------

5.3 - Software preservation White Paper - starting point for discussion

Iolanda Maggio (ESA/Starion) reported [[slides](#)]:

- Software preservation is becoming a bigger topic. The white paper has already been included in the CEOS Work Plan as DATA-24-02.
- The white paper is intended to assist data/software managers with the task of ensuring the long-term preservation of EO missions and data related software, thus improving data accessibility and usability for current and potential future users.
- The intended audience should comprise data and software providers, decision makers and scientists, and data managers/stewards for data centres and repositories.
- Software should be preserved to ensure the maximal value of the data can be retrieved in future. It also provides valuable historical context for scientific data as it reflects the analysis methods and technologies used over time.
- The three main principles are collect, preserve and share, which help to guarantee the robust preservation of software while maximising benefit to, and usage by, potential future users of the software.



- A wide array of challenges face software preservation, from technical complications to legal issues and resource availability.
- A well planned and proactive approach to software preservation may help to avoid some of the most encountered challenges, particularly for larger, well-resourced organisations.
- A dedicated session will be organised at WGISS-59 to share experiences and technical information to be used to draft the White Paper.

Discussion

- Peter Cornillon (University of Rhode Island) asked when developing software, is it done in GitHub, with versioning?
- Iolanda noted that not every agency uses GitHub, and commercial providers likely wouldn't. Commercial providers could be included in this activity, and perhaps invited to WGISS-59.

WGISS-58-14	DSIG to consider inviting external experts to present on software preservation technologies and approaches at WGISS-59.	Due: WGISS-59
--------------------	---	--------------------------

Day 2: Thursday 17 October, 2024

Session 6: STAC Topics

6.1- Introduction to STAC in WGISS

Doug Newman (NASA) reported [[slides](#)]:

- STAC is very amenable to federated access, and will be a good replacement for Open Search.
- There has been wide adoption of STAC throughout the EO community, with a thriving ecosystem.
- WGISS has developed best practices for EO Collection and Granule Discovery with STAC, with the [first release published on GitHub](#) recently.
- NASA is working to replace the Open Search implementation in CWIC with STAC, however there is a need to find some way to link STAC catalog APIs to other agencies.
- Over the last few months of 2024, NASA will work to implement a Collection Search extension for CWIC, ensuring compliance with the best practices.
- The INPE collection metadata will be updated with STAC endpoints
- In 2025, the team plans to begin adding other high-value extensions: Electro-Optical, Projection, View Geometry, Hyperspectral Imagery, Satellite, Query, Sort, etc.
- Doug presented at ESIP 2024 on the WGISS STAC efforts. The primary discussion point was around the concept of 'Super STAC', which encompasses various takes on federated STAC.
- Doug also made a lightning talk on the CWIC federated approach.

Discussion

- Yves Coene (ESA/Solenix) noted the query extension is obsolete, as the community is moving towards a filter extension instead.
- Peter Strobl (EC-JRC) noted that STAC Extensions are growing organically, and there is increasingly a need for good structure to ensure extensions can be distinguished.
- Doug recognised that the STAC community is aware of this weak point, and is working on improving the structure of extensions.

- Brian Terry (SEO) added that the STAC community is currently reviewing the extensions, with some moved to pilot phases. The community has started defining maturity levels to help with these issues.
- Peter noted that STAC is currently very centred on lower level processing up to level 2. CEOS is starting to consider the higher levels as well, where sensor agnostic approaches become important. STAC should think about defining a robust structure which can be applied across diverse archives.
- Doug recognised that in some regard the uptake of STAC was due to fighting against this sort of rigour.
- Sai Kalpana (ISRO) recognised the list of extensions isn't complete, and maybe extensions should be considered in the best practices.
- Tom Sohre (USGS, WGISS Chair) noted that the current document focuses on the discovery aspect, not targeted at the data provider.

WGISS-58-15	DAIG to include additional information related to Data providers (and users experience) in the STAC Best Practices.	Due: WGISS-59
WGISS-58-16	WGISS Secretariat to make sure v1.1 of the Service Discovery Best Practices is online, including on the CEOS Publications page.	Due: ASAP

6.2 - INPE STAC Catalog integration in FedEO

Yves Coene (ESA/Spacebel) reported [[slides](#)]:

- The FedEO team is working to integrate the INPE STAC Catalog into FedEO. This involved a lot of preparatory activities, especially regarding vocabulary.
- FedEO provides OpenSearch and STAC access to federated backend catalogue, independent of partner backend protocols. FedEO supports collection and granule search.
- Collection metadata is ingested into the FedEO collection catalogue, with new INPE collections becoming available automatically.
- Access via alternative protocols and metadata formats is also preserved.
- INPE collections/granules in FedEO are now available via:
 - o STAC API (collection search/collections, two-step search)

- STAC API (cross-collection item search /search)
- Static STAC catalog (by platform, by organisation)
- DIF-10 interface (consumed by CEOS IDN)
- OpenSearch (Collection search)
- OpenSearch (Granule search)
- The team is aware that it will take some time to migrate all datasets to STAC, so will keep the OpenSearch functionality available as well.
- The STAC implementation allows for various searching and browsing methods.
- INPE wishes to organise its entire collection of surface remote sensing data in this framework, including the long term archive.
- STAC backend (granule) catalogue is sufficient to be federated with FedEO (CEOS Connected Data Assets), with a minimum number of changes on the partner side. INPE can also publish additional collections autonomously (as long as keywords are included and exist in thesauri).

Discussion

- Libby Rose (WGISS Secretariat) asked if this changes anything about how FedEO is integrated with the CEOS MIM Database.
- Yves noted there is no client specific for STAC, instead FedEO has a link to the organisation client page, either for OpenSearch or for STAC. This could be also integrated with the CEOS database.

6.3 - CEOS Service Discovery BP - STAC integration

Yves Coene (Spacebel for ESA) reported [\[slides\]](#):

- CEOS Best Practices for Services Metadata and Discovery v1.0 was published in November 2022.
- Extends Best Practices for collection and granule discovery (CEOS OpenSearch Best Practices) to “services” (incl. tools, applications, service endpoints) with intention to “share”.
- Defines core set of metadata properties without imposing single metadata format.
- First step towards sharing tools and applications metadata via CEOS IDN.
- The document has been updated to V1.1 to include STAC and STAC API.

- A number of discovery scenarios are included in the document, including downloadable tools, web GUI tools and coupled services.
- The main changes were added to Chapter 3:
 - o 3.3 Service metadata encoding: STAC was added as an eighth encoding example
 - o 3.5 Service discovery interface: STAC API added.
- No changes were made to the objective of the document and the use cases.

Discussion

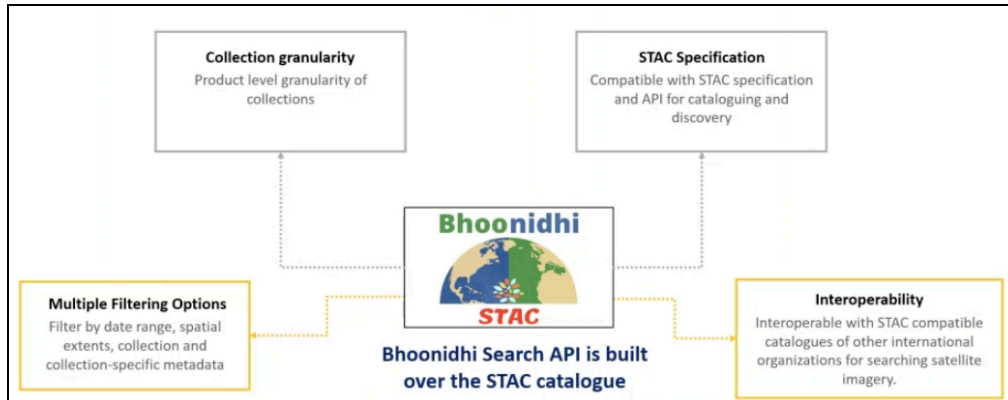
- Brian Terry (SEO) asked whether there are validators for the best practices, in some automated way.
- Yves noted this doesn't yet exist, but the team could look at developing this, noting that the objective of the document was eventually to share metadata via IDN.
- Brian recognised a similar validator tool was successful for STAC.

WGISS-58-17	DAIG to look into a tool to validate a provider STAC interface against the Service Metadata and Discovery Best Practices.	Due: WGISS-59
--------------------	---	--------------------------

6.4 - ISRO STAC overview

T. Sai Kalpana (ISRO) reported [[slides](#)]:

- ISRO is looking at improving the design of their catalogues for better user experience, and have decided to use STAC.
- Developing an accompanying RESTful API to enable search.
- STAC will be integrated into Bhoonidhi, ISRO's data access platform.
- COGs will also be included in the integration, to enable direct cataloguing of data into data cubes.



- The API is used for authentication, search and download.

Discussion

- Doug Newman (NASA) asked how many collections are available via STAC.
- Sai noted it is currently available for 4 satellites, and about 10-12 collections.
- Nitant Dube (ISRO, WGISS Vice-Chair) added that authentication is only needed for download, not discovery. However, some datasets are not open - cal/val, commercial products, etc. There are use cases where you don't want all products to be viewable for anyone.
- Damiano Guerrucci (ESA) noted this is more about authorisation not authentication. However, there is a need to make sure people aren't accessing data through the back-door, if they can discover the datasets this is more likely. This would require a system where a user has to login to see the data.

6.5 - STAC Summit overview and outcomes

Chris Stoner (AWS) reported [[slides](#)]:

- The STAC Summit met before the 2024 ESIP meeting.
- The STAC endpoint should be hosted by a neutral 3rd party, who can act in the best interest of the community. Talking to Radiant Earth at the moment to help keep this running as it scales.
- The STAC Summit was attended by government agencies, partners and collaborators. Discussed about contributing STAC metadata to a federated STAC, and Proof of Concept STAC collections for datasets.
- Many data providers are already creating STAC metadata, or are planning to.

- There was good interest in moving forward with SuperSTAC. A prototype has been created by the North Carolina Institute for Climate Studies (NCICS) with RadiantEarth. However, the work has been paused due to the recent hurricane in the area.
- DevelopmentSeed is continuing to work on STAC Tools Pipelines, which can help data providers get started with STAC.
- Element84 is looking to scale STAC using geoparquet.
- A community leader is needed to drive the SuperSTAC effort.

Discussion

- Brian Terry (SEO) asked which level the SuperSTAC prototype is following.
- Chris noted that the NCICS team is following two - catalog and item. Some performance tests and usability tests. If there are other people who want to try, Chris can connect them with NCICS.
- The team is weighing up the pros and cons of crawling to form a master catalog, or offloading via a cached catalog. The NCICS team started by using the crawling method.
- Doug Newman (NASA) noted that CWIC uses the 'slowest person in the room' method.
- Peter Cornillon (University of Rhode Island) asked about the handling of duplicated datasets.
- Chris noted that the team is going for the easiest approach for our purposes, was going with the easiest approach, and letting the user choose. There is probably another effort to understand duplicates and validate mirrored datasets.
- Peter Strobl (EC-JRC) raised concerns about the large dependency on a single centralised catalog. Having several instances would mitigate this, but would have to make sure they are fully aligned.
- Katie Baynes (NASA) asked about the role of IDN if this SuperSTAC is successful.
- Doug noted that IDN is one of the possible solutions for a centralised collection level catalog.
- Tom Sohre (USGS, WGISS Chair) recognised the team would have to understand how to rank the datasets when returning search results. This might be a role for AI.
- Damiano Guerrucci (ESA) recognised that the governance might be too overwhelming to have all the datasets in a single place. Need to give guidance to users on the best place to go to find the answers they need.

- Tom noted that if the pilot becomes operational, then WGISS has an opportunity to decide whether to point our users to superSTAC, or provide crawlability of the STAC IDN implementation.
- Robert Fletcher (UKSA) noted that the UK Met Office was listed as a data provider, and asked how they are involved. Chris is working with Jeremy Tandell at the Met Office, as AWS houses over 100 PB of open data, including from the Met Office. Have discussed with them about creating a STAC Catalog for their data.

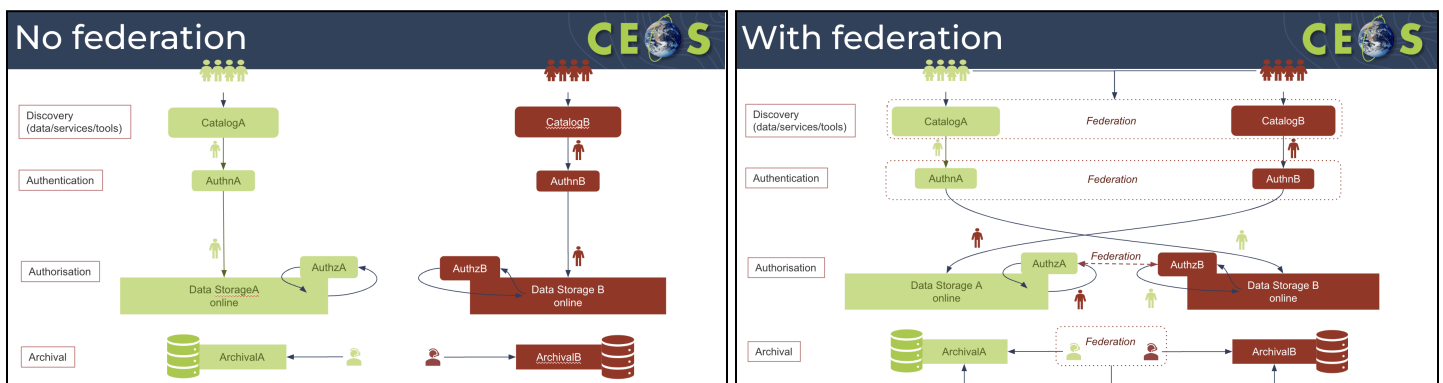
WGISS-58-18	DAIG to connect with the NCICS team developing the SuperSTAC prototype, and remain aware how this interacts with IDN & FedEO.	Due: WGISS-59
--------------------	---	--------------------------

Session 7: Federated Authentication and Authorization

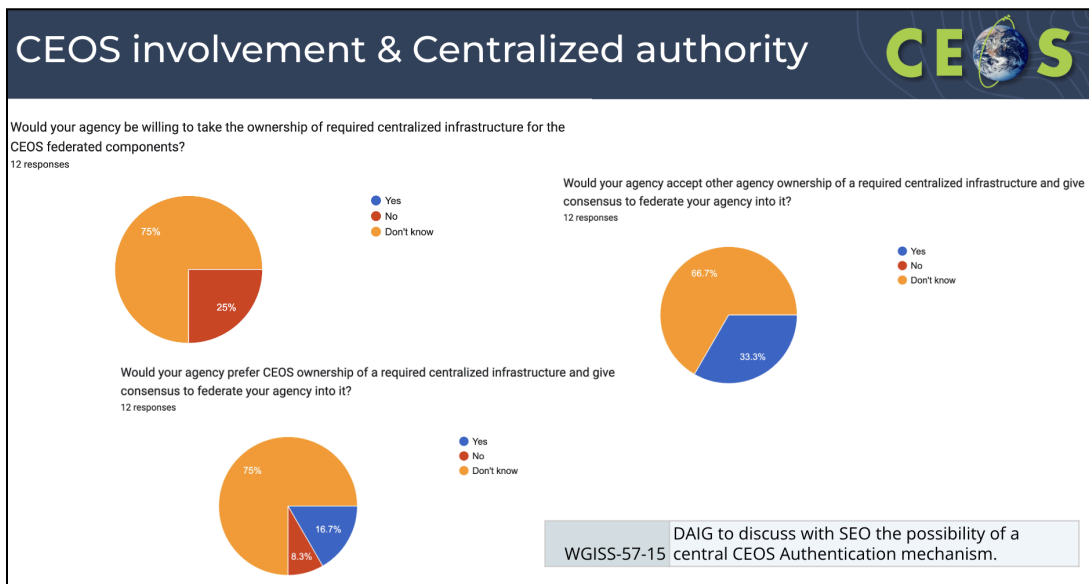
7.1 - Introduction to Federated Authentication and Authorization

Damiano Guerrucci (ESA) and Filippo Marchesi (ESA/Solenix) reported [[slides](#)]:

- A survey was conducted to understand the various interpretations of Federation.
- The following terms are defined in ESA as follows:
 - o **Authentication** is the process of confirming the correctness of the claimed user identity.
 - o **Authorization** is the process of granting or not access to a protected resource to a user, based on information available about such a user.
 - o **Federation** is a group of institutions that agree about policies for exchanging information about their users/data/metadata/process in order to grant access and utilisation of protected resources and services.



- Federation can be done at different levels, including discovery, authentication, authorisation and archival.
- The survey aimed to:
 - o Investigate the current Federated Authentication and Authorization landscape among members
 - o Identify the use cases and scenarios of interest to the group related to this topic
 - o Define steps for the way forward depending on interests emerged
- 12 answers from CEOS Agencies were received. The survey remains open for responses.
- Half of the respondents use a Federated Authentication and Authorisation infrastructure, with another quarter intending to do so in future.
- Interoperability and improved user experience were identified as the main benefits. The largest barriers were around governance and policy.



- There is still a lot of confusion around this topic, and it remains unclear about the way forward.
- There are a number of existing documents, guidelines and standards from across CEOS Agencies.
- Most respondents agreed that a white paper or best practice on this topic would help in the implementation of Federated Authentication and Authorisation with their agency.

Discussion

- Tom Sohre (USGS, WGISS Chair) clarified that this is about user authentication. A white paper could be valuable, but cautioned the group to not spend too much effort where this work has already been done.
- Brian Terry (SEO) asked about the infrastructure requirements on either side for a federated system.
- Filippo noted that Solid is one option, which is an open standard for structuring data, digital identities, and applications on the web.
- Brian recognised that a choice needs to be made as to whether it is centralised or decentralised federated, and it might be good to understand the infrastructure costs for both options.
- Katie Baynes (NASA) suggested understanding why agencies are doing authentication could help understand what solutions would fit best. NASA mostly does it for tracking user communities and understanding their needs.
- Tom recognised that WGISS should consider how surveys are conducted in general. Do we want just one response from each agency?
- Robert Fletcher (UKSA) asked whether the team have considered social media authentication methods? Filippo noted this was included in the survey, and the answers were mostly 'no' or 'it depends'.
- Yousuke Ikehata (JAXA) suggested providing users a choice between Agency authentication accounts, similar to how users can log in to many platforms with either google/facebook/etc.
- Filippo recognised the team isn't yet sure how it would work, and are working on deciding if it is desired. Then, DAIG will ask for support from TEIG and SEO.

WGISS-58-19	STAC Best Practices to include a section about authentication & authorisation, considering the issues regarding discoverability of non-public datasets.	Due: WGISS-59
WGISS-58-20	DAIG to investigate the infrastructure costs of centralised and decentralised authentication methods.	Due: WGISS-59
WGISS-58-21	TEIG to investigate with DAIG and SEO the available technologies for federated authentication, and which would fit with the objectives of CEOS.	Due: WGISS-59

7.2 - Integrity Provenance and Trust - OGC Testbed-20 Report

Yves Coene (Spacebel for ESA) reported [[slides](#)]:

- OGC Testbed-20: Integrity, Provenance and Trust (IPT) uses the Decentralised Identifier (W3C DID) technology, as mentioned in the survey.
- The objectives are to establish Integrity, Provenance, and Trust (IPT) building blocks that are aligned with existing OGC building blocks and adhere to FAIR principles (findable, accessible, interoperable, and reusable). The building blocks should maintain data integrity, track data provenance, and build trust in data across its entire lifecycle.
- EO data supply chain was identified as a use case for the IPT project. The idea is to allow downstream Earth Observation data consumers to “trust” the data, i.e. verify that data consumed is from the original data provider and allow verification of the integrity of the data.
- Will demonstrate infrastructure with credential issuer, credential holder and credential verifier. A subset of the same technology will be combined with the OGC API.
- DID is similar to a DOI (Digital Object Identifier). A DID ‘subject’ is identified by a ‘DID’, for example a EO product. A DID ‘document’ describes the subject and includes an organisation's public key information.
- The Verifiable Credential (W3C VC) can be verified by parties via DID via the Verifiable Presentations (VPs).
- The team has demonstrated feasibility to implement integrity/provenance/trust using W3C DID/VC/VP technology.
- However, implementations/libraries have different limitations: e.g. version of VC, version of JSON-LD, supported key types and encodings, embedded contexts etc.
- Approach and scenarios for IPT (Integrity Provenance and Trust) will be documented in Report OGC 24-033.
- Next steps include integration with OpenID Connect (OIDC) and OAuth.

Discussion

- Robert Fletcher (UKSA) recognised that the principals are similar to those of blockchain. However for this, something would need to be installed on the satellite.

7.4 - Identity Federations at DLR EOC

Ulrich Raape (DLR) reported [[slides](#)]:

- DLR started investigating a User Management System for DLT Earth Observation Centre (EOC) in 2016, called EOC IAM.
- Their approach to account design was initially focused on one account type per protected dataset/system, but wanted to move toward single-sign-on and federation.
- In 2017-2018, DLR demonstrated 'Identity Federation - Deutschland' (IFed-D), as part of an ESA Pathfinder Project.
- The EO-Lab cloud is part of the Strategy Artificial Intelligence of the German Federal Government, and the EOC Geoservice hosts a wide range of free and restricted datasets.
- The EOC UMS acts as 'Service Provider' to the EO-Lab Identity Provider for a set of protected datasets.
- This year, ESA started the EO Exploitation Platform Common Architecture - plus (EOEPCA+), which aims to design and deploy an IAM building block for federated applications/platforms.
- The authorization services are based on OpenPolicyAgent (OPA) and Open Policy Administration Layer (OPAL).
- There are also test and demonstrator activities related to federated discovery and access.
- DLR's terrabyte data analytics platform participates in the German Research Identity Federation, which is used for identity mapping across different organisations.

7.3 - Federated Integration of INPE Datasets

Michael Morahan (NASA/KBR) reported [\[slides\]](#):

- The end goal is to download a granule without authentication (only logging into NASA Earth Data).
- The project demonstrates the collaboration of multi-agency resources INPE STAC, FedEO API, and IDN Search Portal for collection discovery and granule download without authentication/authorization.
- INPE datasets are [available in IDN](#), directly from INPE through STAC. Data can be pulled out directly without authentication.
- This can also be done through FedEO, using OpenSearch.
- The team is in touch with USGS regarding implementing this for Landsat products, and will start testing this soon.

Session 8: Technology Exploration

8.1 - AI/ML White Paper

Yousuke Ikehata (JAXA) reported [[slides](#)]:

- Following WGISS-57, some more examples have been added to Chapters 4 (Initiative and Programs), and Chapter 7 (Data and platforms)
- At WGISS-56, an action was recorded to explore training around AI/ML in cooperation with WGCapD. No suitable topics could be found, but WGCapD shared information about a [training course from EOCollege](#) on this topic.
- WGISS-56-07 called for an inventory of well-known and easy to use training datasets (AI/ML ready data) in the white paper. TEIG has put together a [survey](#) to gather information on what 'AI/ML ready data' means, and will be discussed today.
- A table could be included in the white paper, highlighting which features a dataset should have (mandatory or recommended) to be considered 'AI/ML ready data'.
- AI/ML ready data is separate from CEOS-ARD. A maturity matrix could be considered for this topic.
- Yousuke shared the [ESIP checklist for AI/ML Ready Data](#).
- TEIG would like to postpone the deliverable.

Discussion

- Robert Fletcher (UKSA) asked what types of features a dataset would need to be considered 'AI/ML Ready', in comparison to those required for ARD.
- Tom Sohre (USGS, WGISS Chair) would like to know if the user community wants to package the data differently for use with AI/ML.
- Doug Newman (NASA) noted that RadiantEarth was working on [MLRD \(Machine Learning Ready Data\)](#). The project ran out of funding, but was making some good progress.
- Brian Terry (SEO) noted the architectures for AI/ML processes can vary substantially, but the access patterns don't. What a user is doing once the data is loaded is evolving rapidly.
- Tom suggested WGISS focus on how to provide the biggest value to the user community, noting the rapid pace of developments with AI/ML.

- Brian recognised that ARD requirements would surely overlap with requirements for AI/ML ready data, and perhaps the team should investigate the differences. Having a good, widely available dataset enables more applications.
- Ashutosh Gupta (ISRO) noted the checklist was circulated widely within ISRO, and it was difficult to meet all the criteria. One key difference with ARD is the requirement for labels.
- CEOS-ARD is divided up into product types, which have different requirements. AI/ML requirements should be the same across all products. The CEOS-ARD metadata requirements however may be helpful.
- Libby Rose (WGISS Secretariat) suggested one recommendation for 'AI/ML ready data' is for it to be CEOS-ARD compliant, and any additional AI/ML specifications could be added on top of this.
- Tom suggested harmonising the requirement language with CEOS-ARD: 'threshold' and 'goal'.
- Peter Cornillon (University of Rhode Island) would like a repository of AI/ML datasets somewhere, with links to journal articles which used it, etc.
- Yousuke noted the ESIP checklist was probably put together by users, however the motivations behind the checklist remain unclear. TEIG survey is instead looking at the data provider side.
- Tom raised concerns about postponing the deliverable further, and suggested instead a living document be posted, to reflect the dynamic nature of this sector.
- Brian noted that SEO is developing CEOS-GPT, with the idea being to make the documents more searchable. In testing at the moment, not yet online.

WGISS-58-23	TEIG to consider what additional requirements would be needed on top of CEOS-ARD specifications to make a dataset 'AI/ML ready'.	Due: WGISS-59
WGISS-58-23	WGISS Agencies to answer the TEIG survey about AI/ML ready data .	Due: ASAP
WGISS-58-24	WGISS Secretariat to add a link to the AI/ML GitHub repository to the CEOS website as a 'living document', and invite contributions from the community via comms channels.	Due: ASAP

8.2 - Jupyter Notebooks Best Practices

Yousuke Ikehata (JAXA) reported on behalf of Ester Conway (STFC) [[slides](#)]:

- The best practice has been finished, and is available on google docs.
- Asked how the document should proceed.

Discussion

- Tom Sohre (USGS, WGISS Chair) agreed the document should be distributed as Google Docs to the WGISS community for a two week review period, before being posted online as an endorsed document.
- Regarding future Technology Exploration topics, what should WGISS do next?
- Doug Newman (NASA) noted the AI/ML work was something a bit too broad, where WGISS didn't really have the right experts to contribute.
- Brian Terry (SEO) suggested content identification and provenance, having unique cryptographic identifiers, or distributed storage.
- Tom suggested perhaps something around compression algorithms, to help support issues with growing data volumes.

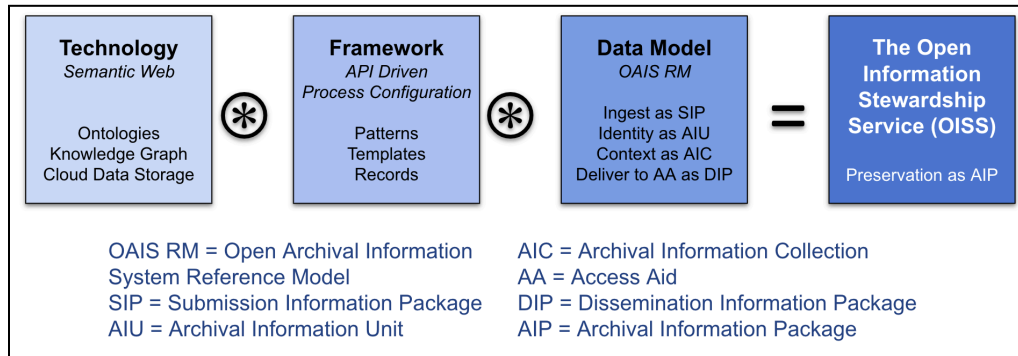
WGISS-58-25	WGISS Secretariat to help TEIG format the Jupyter Notebooks Best Practices into CEOS document template. TEIG should then share the document for review with WGISS members.	Due: ASAP
WGISS-58-26	TEIG to consider starting an activity around compression algorithms (perhaps a white paper), for inclusion in the 2025-2027 CEOS Work Plan.	Due: WGISS-59

8.3 - Knowledge graphs

Ken Casey (NOAA) reported [[slides](#)]:

- Open Information Stewardship Service (OISS) is an example of a system being deployed specifically and explicitly to use knowledge graphs to enable technologies such as AI, Large Language Models (LLM), and Digital Earth Twins.
- OISS is the data workflow manager that NOAA is using to archive and steward data in the cloud, and that NCEI science teams can use, if they choose, to fully support their science workflows.

- Throughout the workflow of data producers to data consumers, each step is written to the knowledge graph. This gives full provenance and transparency - scientific reproducibility.



- Semantic interoperability within a shared reference model is required to allow data to be shared and reused by people and their applications across communities and domains.
- If multiple organisations have similar graphs, a ‘knowledge network’ could be made to connect the different graphs to discover across organisational boundaries.

Discussion

- Doug Newman (NASA) noted that NASA stumbled in not following through with the number association, and not making the tools public. Is the API tool public?
- Ken noted that nothing is public from the NESDIS Common-Cloud Framework (CCF) yet, the team is still waiting for approvals. Graph won't be exposed immediately, but there is an intent to make it public.
- Ryan Berkeimer (NOAA) recognised that NOAA is participating in the open knowledge network with NASA. This needs the interoperability of ontologies and interfaces.
- Ken recognised the human challenges may be harder, enabling uptake, and ensuring data is entered correctly. Enabling people across the spectrum to enhance the information.
- Peter Cornillon (University of Rhode Island) suggested that the system should be configured such that it can interface with large language models (LLMs), to support this new era of searching for information.
- Ken noted that yes, that is the intent. NOAA has to have their own catalog, but also have to make sure it is listed on schema.org. Intent it to make it public in such a way that LLMs can encounter it.

- A simple thing to do would be to inject knowledge graph thinking into the federated catalogs - STAC etc. The CEOS MIM database might be a nice implementation for a knowledge graph - exposing the information as a graph.
- Ryan noted the team is also looking at going backward, configuring the process by talking to it. This could be applied to the Digital Twins (DT) space - a system which maps all the DT requirements to enable federated DT.

WGISS-58-27	Ken Casey to discuss further with the NOAA team about a knowledge graph demonstration using the MIM Database. Once ready, Libby Rose can connect Ken with the MIM Database Team to discuss further.	Due: WGISS-59
--------------------	---	--------------------------

Session 9: Challenges with Data in the Cloud

9.1 - Review of CGMS Best Practices

Matt Paget (CSIRO) reported [\[slides\]](#):

- The Coordination Group for Meteorological Satellites (CGMS) was invited to present at WGISS-56 on the Cloud Best Practices and Approach to International Collaboration.
- How we use this document going forward is for discussion today.
- The document is largely from an agency perspective about how to go about transitioning to the cloud, and the infrastructure and code requirements. Governance and financial challenges are discussed as well.
- The migration strategies table is a helpful reference for all practitioners as we progressively include cloud-computing in more of our new and legacy workflows.
- Usefulness of the document will be dependent on where an Organisation is on its cloud journey.
- The document remains neutral with respect to final solutions mostly applied, and is not specific to meteorology.
- Perhaps the document is too generic, but unclear about the way forward here.

Discussion

- Tom Sohre (USGS, WGISS Chair) what should WGISS do with this document, if anything? The document is linked from the slides, which are available on the CEOS website, and perhaps that is specific enough. As a peer group, we have provided our review and comments.

WGISS-58-28	Robert Fletcher (and others interested) to review the CGMS Cloud Best Practices document. DAIG to compile comments and share the feedback with CGMS on behalf of CEOS.	Due: Q4 2024
--------------------	--	-------------------------

9.2 - Pain points identified by NASA

Doug Newman (NASA) reported [\[slides\]](#):

- NASA made some early decisions around the usage of S3, which is causing problems for the use of S3 for data, logs, serverless applications and AI/ML pipelines. To mitigate this, agencies should be adaptable.
- On the cloud, the data provider doesn't control the configuration of the platform. NASA had issues with the user metrics methods negatively impacting the user experience. The solution for this isn't yet clear.
- There are cost issues regarding unbound egress. However, one has to be careful that the solution costs less than the risk.
- Tempting to take everything from on-premises and dump it on the cloud, however providers should make sure the data is optimised for cloud access. Archive formats perform poorly in a cloud environment. Data providers should leverage managed services by cloud providers.
- Vendor lock-in is a real issue, but the benefits can outway the risks. To totally mitigate vendor lock-in risks, use only EC2/AZM/GCE, EKS and S3/ABS/Google Storage.

Discussion

- Brian Terry (SEO) asked for Doug's opinion of what is the highest value managed service?
- Doug Newman (NASA) the EC2Spon market has reduced compute costs for on demand processes by 70%. However, there is a risk that the instance might be ripped away, but a very small percentage of the time.
- NASA is still transitioning to the cloud - the end of 2026 is the deadline to get all data and services up to the cloud. On schedule for the data side.
- NASA is looking at multi cloud vendor distribution. Currently with AWS as they had what was needed in 2016.
- Peter Cornillion (University of Rhode Island) noted sidecar solutions, where metadata is stored separately.

- Katie Baynes (NASA) noted that once data is moved to the cloud, providers then have to work to change access patterns. This is a cultural change, which is slow to happen.
- Doug recognised that on-premises download and compute will be unsustainable with the bigger data products coming down the line. No longer cost effective to do the processing on your own hardware.
- Tom Maisperger (USGS) is unsure whether USGS has made the right investment in capacity development. There are around 100 thousand Landsat users every year, most of which are not big data users.
- Brian Terry (SEO) noted this is partly why SEO is offering the CEOS Analytics Lab, and doing some training as well. Allows for cloud native processing - at different scales. The higher levels of scalability haven't yet been opened up.
- Tom Sohre (USGS, WGISS Chair) also noted the issues with data replication issues, as organisations often want a copy of data.
- Alex Leith (Auspatious) recognised that Planet once had a massive investment from Google, and shifted all data from AWS to Google, in a relatively short amount of time. It might be easier than once thought to shift between cloud providers.

9.3 - Digital Earth experiences

Alex Leith (Auspatious) reported [[slides](#)]:

- Data formats are really important, due to the latency and scalability.
- Digital Earth Australia started in 2010, and had to digitise data, process the data, develop the cloud and make it available online.
- For Digital Earth Africa, there was an emphasis on making decision-ready products. There was a fair bit of work to organise the data before productivity could begin.
- For Digital Earth Pacific, the data isn't downloaded, and taken straight from Microsoft Planetary Computer, with STAC APIs and Python libraries.

Capability	Digital Earth Australia (2010)	Digital Earth Africa (2020)	Digital Earth Pacific (2023)
Landsat archive fully accessible	✗	✓	✓
Analysis Ready Data definition available	✗	✓	✓
Landsat level 2 data available	✗	✓	✓
Sentinel-2 level 2 data available	✗	✓	✓
Landsat and S-2 accessible in the Cloud	✗	✗	✓
Sentinel-1 level 2 (RTC) available	✗	✗	✓
STAC for data access widely supported	✗	✗	✓

- A third party makes Sentinel-2 COGs available.
- Common STAC metadata for all the three major data products (Sentinel-1 & -2, Landsat) is a game changer.
- The global coastlines product from Digital Earth Australia was originally developed for a supercomputer and file systems. The process was redeveloped to load data directly from S3 directly into memory and write directly back to object storage.
- With Digital Earth Africa, there was a mistake made with an infinite loop, which was a costly mistake.
- Another mistake was made around the security of the network, when the direct connection to S3 (VPC Endpoint) was turned off and 400 TB of data was pushed across the NAT gateway.
- The processing of Landsat Collection 2 wouldn't have been possible without the cloud. The scale possible for the cost is incredible.

Session 10: Closing Session

10.1 - WGISS-58 Action Review

Libby Rose (WGISS Secretariat) reported:

- Reviewed the actions recorded during WGISS-58.

WGISS-58-29	WGISS members should ensure they have a GitHub account, and are a member of the CEOS Organisational account . Contact Brian Terry (SEO) for assistance if needed.	Due: WGISS-59
--------------------	---	--------------------------

WGISS-58-30	WGISS Secretariat to put together guidance for Interest Groups about surveying CEOS Agencies on various topics.	Due: WGISS-59
--------------------	---	--------------------------

10.2 - Closing Remarks

Tom Sohre (USGS, WGISS Chair) reported [[slides](#)]:

- WGISS-59 will be held in Bangkok, Thailand, hosted by GISTDA, on March 24-27 2025.
- WGISS-60 will be hosted by DLR in Oberpfaffenhofen, on October 13-17. WGISS members are asked to check for conflicts with these dates.
- Nominations for WGISS Vice-Chair for 2026-2027, followed by Chair for 2028-2029 can be sent to Tom.
- Thanked all for participating and contributing to the discussions.

Appendix A: List of Participants

In-person participants

Affiliation	Name
CEOS Executive Officer	Steven Ramage
CNES	Fournier Hugo
ESA	Mirko Albani
ESA	Damiano Guerrucci
ESA	Yves Coene
ESA/Solenix	Filippo Marchesi
ESA/Starion	Iolanda Maggio
European Commission	Peter Strobl
Institute of Geographic Sciences and Natural Resources Research, CAS	Shi Ruixiang
Institute of Geographic Sciences and Natural Resources Research, CAS	Li Limin
ISRO NRSC	Sai Kalpana
JAXA	Yousuke Ikehata
NASA	Doug Newman
NASA	Katie Baynes
NASA/KBR	Michael Morahan
SEO	Brian Terry
UKSA	Robert Fletcher
University of Rhode Island	Peter Cornillon
USGS	Tom Sohre
USGS	Timothy Stryker
USGS	Ryan Longhenry
USGS/Aerospace	Steve Covington
USGS/KBR	Megan Rush
USGS/KBR	Christopher Barnes
WGISS Secretariat	Libby Rose

Virtual Participants

Affiliation	Name
Auspatious	Alex Leith
AWS	Chris Stoner
CSA	Joey Martin
CSIRO	Matt Paget
DLR	Jonas Eberle
DLR	Katrin Molch
DLR	Ulrich Raape
DLR	Felix Feckler
Geoscience Australia	Simon Oliver
GEOSEC	Paola De Salvo
ISRO	Nitant Dube
ISRO	Ashush Gupta
JAXA	Makoto Natsuisaka
NOAA	Sarah Menassian
NOAA	Ken Casey
NOAA	Rich Baldwin
NOAA	Ryan Berkheimer
University of Bern	Stefan Wunderle
USGS	Tom Maiersperger
USGS	Christopher Torbert
	Philip Jones

Appendix B: Actions

WGISS-58-01	Interest Group leads to review the GEO Open Data Open Knowledge (ODOK) actions to identify opportunities for collaboration.	Due: Q4 2024
WGISS-58-02	Iolanda Maggio to circulate the latest version of the Collections Management document to the writing team, WGISS Exec & WGISS-58 participants for final review.	Due: 31st October
WGISS-58-03	DSIG to update the document governance process to define that approval by CEOS Plenary is only necessary for high-level strategic documents, especially where it is a collaborative effort across CEOS, or otherwise where deemed appropriate when submitted to work plan. This should then be brought to CEOS SEC by the WGISS Chair, to ensure it aligns with broader WGISS processes.	Due: 15 November 2024
WGISS-58-04	WGISS Secretariat to discuss with the SEO about increasing the impact of information posted on the CEOS website.	Due: WGISS-59
WGISS-58-05	In developing the 2025-2027 CEOS Work Plan, Interest Group leads should identify the target audience(s) for their deliverable to inform the communication approach.	Due: Q1 2024
WGISS-58-06	Mirko Albani to contact ISRO about the processing of their AVHRR data and contribution to the AVHRR recovery activities under DSIG lead.	Due: 15 November 2024
WGISS-58-07	Mirko Albani to connect with SST-VC regarding presenting the DSIG AVHRR work at GHRSS, including for potential future activities on the heritage data recovery topic.	Due: WGISS-59
WGISS-58-08	Matt Paget to share information with DSIG about the formats used by CSIRO for AVHRR data.	Due: 15 November 2024
WGISS-58-09	Matt Paget & Stefan Wunderle discuss if/how the two AVHRR projects can converge, including on the re-calibration of data.	Due: WGISS-59
WGISS-58-10	Anyone interested in using CoMET once it becomes available to non-NOAA users should send Sarah Menassian (NOAA) an email.	Due: WGISS-59

WGISS-58-11	DAIG to consider whether a CEOS tool should be created to develop a collection metadata file (ISO/STAC), similar to CoMET with maturity matrix support. Consider how this will connect with FedEO & IDN.	Due: WGISS-59
WGISS-58-12	NOAA CoMET team to discuss whether the tool will be open source, and consider how CEOS could build on this work to make a version applicable for CEOS.	Due: WGISS-60
WGISS-58-13	WGISS Secretariat to come up with some language around some related resources for WGISS, and share with Exec to suggest which links should be included.	Due: WGISS-59
WGISS-58-14	DSIG to consider inviting external experts to present on software preservation technologies and approaches at WGISS-59.	Due: WGISS-59
WGISS-58-15	DAIG to include additional information related to Data providers (and users experience) in the STAC Best Practices.	Due: WGISS-59
WGISS-58-16	WGISS Secretariat to make sure v1.1 of the Service Discovery Best Practices is online, including on the CEOS Publications page.	Due: ASAP
WGISS-58-17	DAIG to look into a tool to validate a provider STAC interface against the Service Metadata and Discovery Best Practices.	Due: WGISS-59
WGISS-58-18	DAIG to connect with the NCICS team developing the SuperSTAC prototype, and remain aware how this interacts with IDN & FedEO.	Due: WGISS-59
WGISS-58-19	STAC Best Practices to include a section about authentication & authorisation, considering the issues regarding discoverability of non-public datasets.	Due: WGISS-59
WGISS-58-20	DAIG to investigate the infrastructure costs of centralised and decentralised authentication methods.	Due: WGISS-59
WGISS-58-21	TEIG to investigate with DAIG and SEO the available technologies for federated authentication, and which would fit with the objectives of CEOS.	Due: WGISS-59
WGISS-58-22	TEIG to consider what additional requirements would be	Due:

	needed on top of CEOS-ARD specifications to make a dataset 'AI/ML ready'.	WGISS-59
WGISS-58-23	WGISS Agencies to answer the TEIG survey about AI/ML ready data .	Due: ASAP
WGISS-58-24	WGISS Secretariat to add a link to the AI/ML GitHub repository to the CEOS website as a 'living document', and invite contributions from the community via comms channels.	Due: ASAP
WGISS-58-25	WGISS Secretariat to help TEIG format the Jupyter Notebooks Best Practices into CEOS document template. TEIG should then share the document for review with WGISS members.	Due: ASAP
WGISS-58-26	TEIG to consider starting an activity around compression algorithms (perhaps a white paper), for inclusion in the 2025-2027 CEOS Work Plan.	Due: WGISS-59
WGISS-58-27	Ken Casey to discuss further with the NOAA team about a knowledge graph demonstration using the MIM Database. Once ready, Libby Rose can connect Ken with the MIM Database Team to discuss further.	Due: WGISS-59
WGISS-58-28	Robert Fletcher (and others interested) to review the CGMS Cloud Best Practices document. DAIG to compile comments and share the feedback with CGMS on behalf of CEOS.	Due: Q4 2024
WGISS-58-29	WGISS members should ensure they have a GitHub account, and are a member of the CEOS Organisational account . Contact Brian Terry (SEO) for assistance if needed.	Due: WGISS-59
WGISS-58-30	WGISS Secretariat to put together guidance for Interest Groups about surveying CEOS Agencies on various topics.	Due: WGISS-59