Minutes of the 57th Meeting of the CEOS Working Group on Information Systems & Services (WGISS)

Chaired by USGS
Sydney, Australia

Table of Contents

Day 1: Monday 4th February, 2024

Session 1: Opening Session

1.1 - Welcome and Review of Agenda

Tom Sohre (USGS, WGISS Chair) welcomed all to WGISS-57, and thanked Geoscience Australia and CSIRO for hosting the meeting. All participants were invited to introduce themselves. The full list of participants can be found in Appendix A.

Tom reviewed the agenda for WGISS-57. The meeting is four days for WGISS-57, rather than the three days for previous meetings, to allow for plenty of time for discussions.

1.2 - Welcome from Geoscience Australia

Alison Rose (Geoscience Australia) reported:

− Alison Rose is the Chief of the Space Division at Geoscience Australia. The division has two clear remits: positional and navigation, and Earth observation.

− Welcomed all to Sydney and Australia for WGISS-57, and wished the group well for a productive meeting. Alison thanked CSIRO for providing the catering over the course of the week.

− Acknowledged the traditional custodians of the land on which we meet today, and paid respects to elders past, present and emerging.

− The GA Vision is to be a world leading organisation informing evidence-based decisions through integrated earth sciences to secure Australia's future.

− A number of key programmes across GA:

  ○ Positioning Australia

  ○ Exploring for the Future

  ○ Locate: Australian Climate Services and Digital Atlas of Australia

  ○ Digital Earth Australia

  ○ Critical Minerals

− All five programmes are about location and are strongly supported by the space based programme.
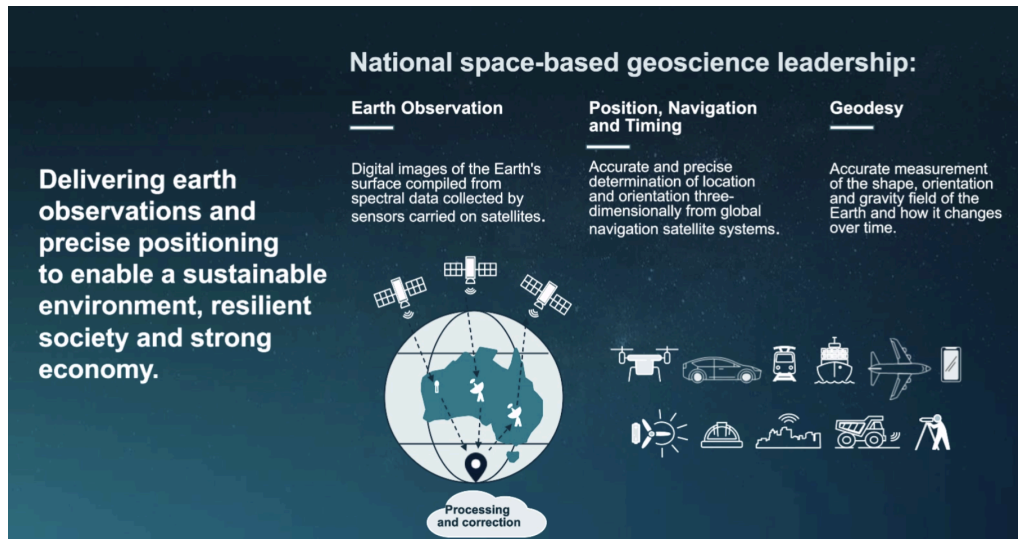
- GA is a key partner in the EO from Space Roadmap, Australian Civil Space Strategy and the Bushfire EO Taskforce.

- The 2028 Strategy aims to achieve:
  - 10cm positioning across Australia and 3-5 cm in areas of mobile coverage
  - EO platform for government and business
  - Assuring streamlined access to satellite land imaging for the nation.

- Positioning Australia Branch is the national authority on geodesy and position verification.

- GA leads the Alice Springs Ground Station Facility. Four antennas which support Landsat missions with data downlink and command/control.

- Promote the collection of land imaging data for Australian user needs.

- GA works closely with colleagues in USGS and the European Commission on assuring access to data, and supports 25 nations in South-East Asia and the Pacific on access to satellite data for their regions.

- Data is made available to the Digital Earth branch, which provides continental-scale time series of medium resolution analysis ready Earth observation data and derivative products.

- End use is really important, as satellite data brings A$5.3B per annum to the Australian economy, with projections showing this could grow to A$2T by 2030.

- Digital Earth Australia covers the below applications across Australia:
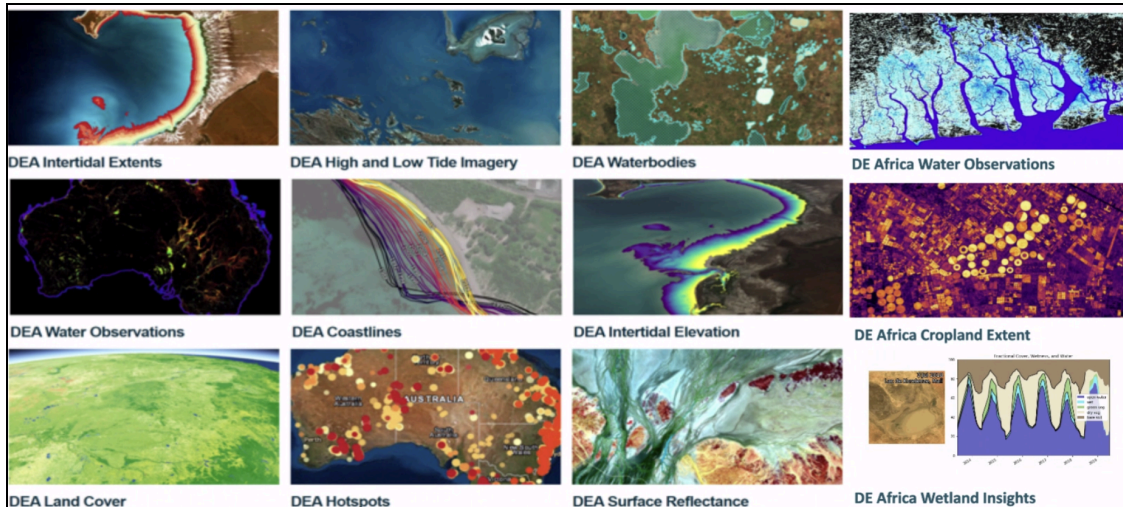
- The data was a key input into the Australian State of the Climate report published in 2021.

- The focus is on bridging upstream and downstream sectors, integrating all types of space based data and analysis to provide products to end users.

- The Digital Atlas of Australia is an interactive, secure and easy to use geospatial platform. Explore, analyse and visualise data by location, and fuse multiple publicly available datasets, as well as the option to add additional datasets. The platform is in Beta phase at the moment.

- The work of WGISS helps to ensure the most benefit from satellite data. GA is excited about the potential of AI & ML, and looks forward to how WGISS explores this topic. The interoperability framework is also a key item of WGISS work for GA.

- GA looks forward to contributing to WGISS alongside other Australian Government partners.

Discussion

- Tom Sohre (USGS, WGISS Chair) noted many of the topics mentioned by Alison will be discussed over the next couple of days, and thanked GA for hosting the meeting.

- Steve Covington (USGS) resonated with Alison's message regarding the value of CEOS to help tackle the challenges with EO data, and make best use of the data.

- Nitant Dube (ISRO, WGISS Vice-Chair) asked about the coastal change hotspot product, and where it is hosted. Alison noted it was built on the Digital Earth Australia platform, but has also been implemented on the Digital Earth Africa platform as well.

1.3 - WGISS Chair Report

Tom Sohre (USGS, WGISS Chair) reported:

- Tom transitioned to WGISS Chair at CEOS Plenary last year. His term as Chair has been relatively short, with this meeting being earlier than usual.

- The WGISS Terms of Reference state the vision is:

  - To enable Earth observation data and information to be more accessible and usable to both data providers and data users world-wide through international coordination.

  - To foster easier exchange of Earth observation and related data and information to meet the requirements of users and data providers.

  - To foster the development of best practices and encourage the development of interoperable services that exploit space-borne Earth observation data.

  - To enhance the complementarity, interoperability and standardisation of Earth observation data and information management and services with other types of geospatial data such as in situ data.

- WGISS has four Interest Groups:

  - Data Preservation and Stewardship

  - Discovery and Access

  - Interoperability and Use

  - Technology Exploration

- The Interoperability and Use Interest Group intends to have members from across CEOS as the issue is a broad topic covering many different aspects. At WGISS-55, it was decided the group would be led by the WGISS Vice-Chair, to help develop those relationships across CEOS. The lead role will transition with the transition of the Vice-Chair role.

- The System Engineering Office (SEO) is a key partner for WGISS, but not necessarily part of the WGISS organisational structure.

- WGISS members are asked to become involved in one or more Interest Groups. Active members are asked to consider their respective organisations and whether there are individuals who would benefit from being involved.

- WGISS is also looking for nominations for Vice Chair (Nov 2025 - 2027) and subsequently Chair (Nov 2027 - 2029), as well as hosts for upcoming WGISS meetings (WGISS-59 in Spring 2025 or WGISS-60 in Fall 2025).

- Highlighted some upcoming Working Group and other CEOS meetings. Most are open for remote participation, and details can be found [here](). LSI-VC-15 will be one in particular to watch, as they are hosting an industry engagement session with local Japanese companies.

- CEOS Plenary welcomed the Canadian Space Agency (CSA) as 2024 CEOS Chair. CSA will prioritise one headline theme of biodiversity during their Chair term. CEOS Plenary also welcomed the Japan Aerospace Exploration Agency (JAXA) as SIT Chair for the two-year term of 2024-2025.

- CEOS Plenary also endorsed the following items:

  ○ Nomination of United Kingdom Space Agency (UKSA) as 2025 CEOS Chair in representation of the Europe/Africa region

  ○ Nomination of the National Aeronautics and Space Administration (NASA) as SIT Vice Chair for 2024-2025 and SIT Chair for 2026-2027

  ○ Dr. Nitant Dube of ISRO as WGISS Vice Chair for two years (2024-2025), followed by WGISS Chair for two years (2026-2027)

  ○ The Ecosystem Extent Task Team white paper and recommendations

  ○ The CEOS Interoperability Framework and Roadmap

  ○ The 2023 CEOS Communications Strategy

  ○ The New Space Task Team white paper and recommendations

Session 2: WGISS Strategic Planning

2.2 - WGISS-56 Action Review

Libby Rose (WGISS Secretariat) reported:

- One action from WGISS-52 in October 2021 and one from WGISS-54 in October 2022 remain open and in progress.

- Three actions from WGISS-55 in April last year remain open and in progress.

- WGISS-56 actions are progressing well, with 13 closed to date, and 5 in progress. Updates about many of those actions will be presented and discussed in the coming days.

- 12 actions from WGISS-56 are open, with no progress yet recorded. There are three in particular which have been picked out to discuss today on the way forward.

- WGISS-56-01: *WGISS to document cloud-enabled dataset recommendations including how data providers avoid data replication.* In reviewing the minutes for context, this action was recorded following CSIRO's agency report, and discussions around the EASI platform.

- WGISS-56-10: *WGISS to review the GEO Data Policy to understand how that might impact CEOS.* This action was recorded following the SEO report, where the GIDTT was mentioned.

- WGISS-56-25: *WGISS to review CGMS Cloud Best practices document and evaluate whether WGISS should promote it and/or develop a CEOS Cloud Best Practices document.* This action followed the presentation from CGMS representatives, where the desire to collaborate was expressed. The minutes note that Nitant suggested WGISS provide feedback to CGMS on the document.

- No clear assignees were recorded for these three actions, and it should be discussed today how to take these topics forward.

Discussion

- Regarding the CGMS Cloud Practices Document (WGISS-56-25), Nitant Dube (ISRO, WGISS Vice Chair) is a co-author on that document. The document was shared with WGISS ahead of WGISS-56.

- CGMS is also adapting some of the CEOS best practices. Should we adapt the CGMS document, or develop our own?

- Mirko Albani (ESA) noted it would be a good approach for us to review and consider whether the document is good for WGISS.

- Tom Sohre (USGS, WGISS Chair) recognised there is not a great place to share documents which are not our own. If we have member documents should we be sharing them?

- Mirko noted a discussion is scheduled for Wednesday on the reorganisation of the website, including the documents page.

WGISS-57-01	Libby to work with Exec to find a place on the website for external documents as well as a process to propose new linked documents and review existing links for continuing relevance.	**Due: ASAP**

- Should WGISS-56-25 be assigned to the Data Access Interest Group?

- Regarding the GEO topics (WGISS-56-10), Mirko will discuss this with the SEO, including about the GIDTT. Richard Moreno (CNES) is the Chair of the Data Working Group in GEO. Not sure what the data policy document is referring to.

- GIDTT is working on the infrastructure of GEOSS activity. A number of options are being evaluated. There are a number of subgroups of the GIDTT, including one on space. Thinking about having a more searchable aspect of the data catalogue.

- Not sure what the 'Data Policy' document is referring to. GEO does a lot for open data.

WGISS-57-02	Mirko, Richard, and Dave Borges (SEO) to coordinate with GEO POC (Paola De Salvo) to provide GEO data topics (incl. GIDTT status) presentation at WGISS-58.	Due: WGISS-58

- Tom Sohre (USGS, WGISS Chair) noted that it is not clear who the points of contact (POC) for WGISS at peer organisations are, such as GEO. If we want to have regular contact with them, it would be nice to have a POC.

- Richard Moreno (CNES) noted it would be nice to have a CEOS WGISS cloud best practices. CGMS is likely looking more at models rather than observations. Computation power is very important for CGMS. Would not expect the CGMS document to fit CEOS purposes.

- Where does this fit within our work planning and prioritising? Likely within DAIG but lots of other activities to also consider there.

- Doug Newman (NASA) recognised it is good to reference an external document where we don't have our own - but should also consider endorsing it.

- Matt Paget (CSIRO) noted that a number of agencies have moved to using the cloud over the last five years. It is still a rather fast moving space, so rather than just producing a document for documents sake, it is worth considering the audience. Less need for a cloud best practice for those already active in the space. But WGISS should demonstrate the experiences of those already doing it to those who aren't.

- Steve Covington (USGS) asked whether there is another approach to tackle this topic, e.g. simplifying to "Best practices for enabling algorithm to cloud exploitation". Maybe the broad topic of "cloud best practices" is too big, as there are different needs for the cloud.

– Tom suggested that instead of publishing papers, would it be best to take information from agencies about lessons learned - and give a way for other agencies to 'vote' on the topic? If there is a common practice between the WGISS agencies.

– A repository approach could be good to streamline the process.

– Libby Rose (WGISS Secretariat) noted the SEO is setting up a CEOS organisational account on GitHub. Doug added that the STAC best practices have been put on GitHub.

– Nitant Dube (ISRO, WGISS Vice-Chair) questioned whether opening up the document could cause too many people to contribute. CEOS would still be the moderator.

– Alex Leith (SEO/Auspatious) noted that this approach is how STAC works.

WGISS-57-03	WGISS to demonstrate the use of a moderated repo approach (GitHub) for Best Practices (or other document) development/feedback. Libby to coordinate with SEO to identify processes (and training) for WGISS staff on CEOS repository. Libby/Doug to coordinate with SEO re: current WGISS github account.	**Due: ASAP**

– Steve noted that looking USGS is looking the other direction - how to provide certification on the copies of the USGS Landsat products. The data is being propagated in an unmonitored way.

– From Alex's experience in using Planetary computer's Landsat products - they are not as good as the USGS copies. It would be good to have a certification process.

– Doug Newman (NASA) recognised this is a valid topic, but not sure if WGISS is in a position to present any recommendations on this topic at the moment.

– Incentive should be the user's confidence in the scientific data, however it has been found that the people who use Google Earth Engine don't care.

– The Sentinel-2 archive in AWS is not perfect, but it being available in COG is worth it. The other option is not doing the research.

– WGISS could come up with a methodology to assess and validate a replicate copy of data and publish it. Have a 'report card', with demonstrable and clear methods. Results would be published to hopefully incentivise the big companies to produce the best product.

- There are many other satellite products which are available on the cloud, but not necessarily cloud enabled. Sentinel-1, for example, requires some processing to get to a scientific product.

- WGISS-56-01 should be closed and replaced with a new action.

- Alex noted there are some tools which can be taken advantage of to maintain a collection to enable replication.

WGISS-57-04	WGISS Exec to convene a small group to tackle the topics around data replication in the cloud, using the CGMS Cloud Best practices document as a reference. Start with a GitHub repository and sketch an outline. Alex Leith and Matt Paget will join the group.	**Due: Q2 2024**

2.1 - WGISS Work Plan Status

Tom Sohre (USGS, WGISS Chair) reported:

- The CEOS Strategic Guidance document, written in 2013, gives an overview of the long term goals for CES broadly.

- Working Groups are tasked with preparing Terms of Reference, which may be supported by a short Working Group Work Plan. WGISS did previously write a specific WGISS work plan up to 2018, which went into more detail on how they will accomplish the tasks listed in the higher level CEOS Work Plan.

- WGISS Objectives falls into three broad categories: support for CEOS, addressing issues related to data management, access and use, and knowledge exchange.

- Typically WGISS has had 1-4 activities per year entered into the work plan. In most cases a document was produced as an output.

- Currently have 6 deliverables open within the work plan.

- Topics for consideration at WGISS-57:

  ○ Collaborate with SEO on CEOS Analytics Lab

  ○ Progress on implementation of Interoperability Framework: Handbook, Demonstrators, Maturity

  ○ Support for Biodiversity: Data Cubes, AI/ML, Interoperability

  ○ Review of past "Info Sharing" (i.e. Open Science, Open Data Cubes, Digital Twins) -> relevant topics for WGISS?

- New Space Task Team: How to Incorporate into WGISS?

- Resourcing our work (Recruiting)

- Update WGISS Outreach Info

- Identify ways to improve feedback for Best Practices

- Identify process improvements for "knowledge sharing"

- Develop multi-year WGISS Work Plan

- What is the priority of these items? There are a lot of things WGISS could do, but need to be realistic about resources and prioritise the activities accordingly.

- Lots of information-sharing presentations in the last few years, what were the goals of this? Is it just about information sharing, or is it an area for WGISS to engage?

Session 3: Agency Reports

**3.1 - Geoscience Australia**

Simon Oliver (GA) reported:

- Thanked Alison for providing an overview of GA earlier this morning.

- GA covers Minerals, Energy, Groundwater, Positioning, Location Information, Oceans, Reefs, Antarctica, Community Safety and Earth Observation.

- Work with satellites started with the Australian Landsat Station established in 1979, in Alice Springs.

- Analysis Ready Data has been a focus area for GA for a number of years, with GA fundamental in establishing the CEOS Analysis Ready Data (CEOS-ARD) framework.

- A project on surface reflectance equivalence across different providers is in development. Like bands from different sensors don't currently align well across different providers.

- GA produces their own products from Landsat and Sentinel Level 1 data. There are equivalent pipelines for Landsat and Sentinel-2, including versioning of products and datasets.

- Within the GA explorer site, a user can explore the original and derivative metadata for each dataset.

- Digital Earth Australia (DEA) Knowledge Hub provides everything a user needs to know about DEA products and services. Includes technology alerts and change logs.

− Digital Earth Australia Sandbox is the jupyter notebook environment for users to play with satellite data.

− Open repository, open for contributions from others as well.

− All data and collections available as STAC v1.0.0 Items, Catalogues and Collections.

− GA was one of the primary developers of the Open Data Cube (ODC) project. The project is now managed by an independent multi-organisation steering council. A new major release (v2.0) is planned soon. Open Source Geospatial Council (OS GEO) recently recognised ODC as a foundational community resource.

− CSIRO, NASA and USGS also contributed heavily to the ODC project as well.

− Responding to rasters and common data usage. GA would like to do more with hyperspectral and lidar data, and enable searching by spectral range.

− Copernicus Australasia Regional Data Hub is working on a major update. The data hub provides a local copy of Sentinel data, and the team is working on switching to the Copernicus Data Space Ecosystem framework.

3.2 - NOAA

Ken Casey (NOAA) reported:

− NOAA Commercial Data Program engages with the commercial sector through pilots and acquisition of operational satellite data-as-a-service for commercial data to help improve weather forecasts and provide risk reduction to the overall observing system. Currently two lines of effort: Commercial Purchase - Radio Occultation Data Buy, and Commercial Weather Data Pilot (CWDP) Round 4 - Ocean Surface Winds.

− Industry Proving Grounds (IPG) is an $85 million investment to promote the development and use of actionable climate information, led by NCEI. Focused on three industries (Finance and reinsurance, retail and architecture, and engineering) to inform sector-relevant data tools and user-friendly climate data.

− The NESDIS Common Cloud Framework (NCCF) is a framework of reusable services providing data ingest, product generation, science development, data access, and long term data preservation and stewardship.

− Developing a new platform called the Open Information Stewardship Service (OISS), which will leverage NCCF services and knowledge graphs (KG) in a semantically-enabled, linked open data interoperability framework to enable full provenance in a scalable, repeatable, and transparent platform.

NESDIS Common Cloud Framework

The NCCF Open Information Stewardship Service (OISS)
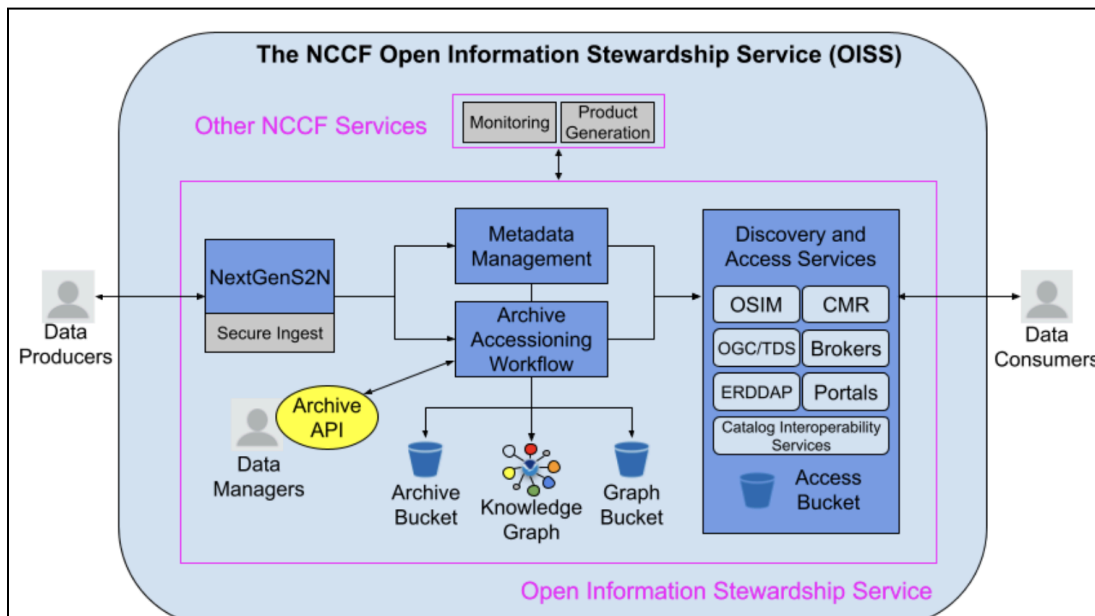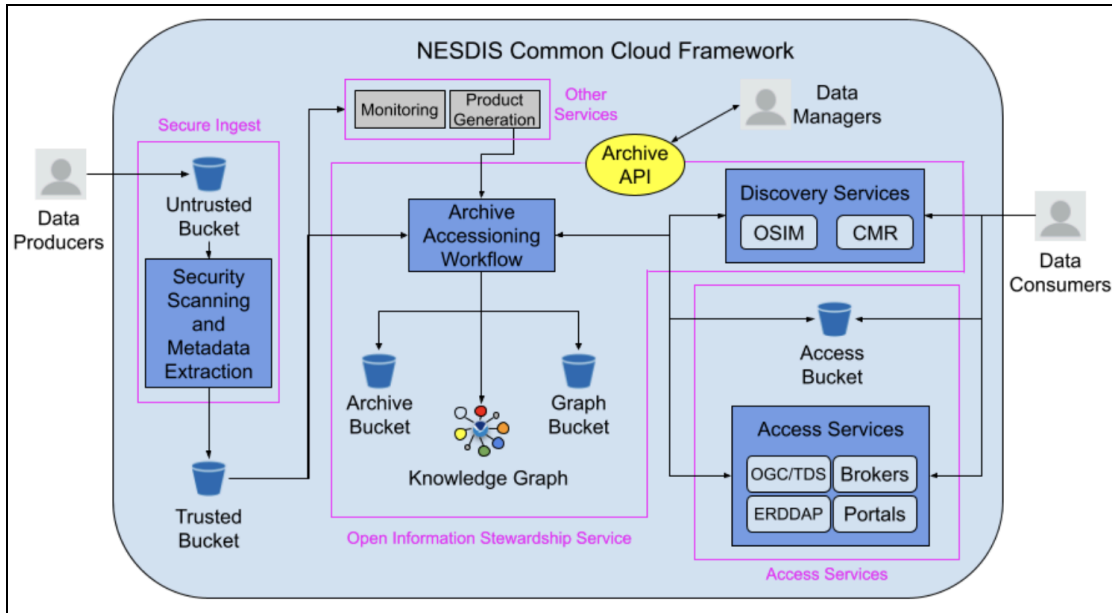
- OISS will allow users to create processes in the event-driven framework, execute AIU (Archival Information Unit), AIC (Archival Information Collection), and DIP (Dissemination Information Package) tasks, move data to Archive and Access buckets, and write results to the Graph Bucket and the KG.

- Fully liked JSON-LD products are useful for Web 3.0 clients (e.g. generative AI) for developing large language models (LLMs).

- Semantic interoperability within a shared reference model is required to allow data to be shared and reused across communities and domains.

- The power of a fully integrated knowledge graph (KG) includes:

  ○ Every archive workflow step is tracked and considered "data", enabling scientific reproducibility and transparency

  ○ Semantic interoperability allows for the KG to be shared across people and apps

  ○ Linkages to other concepts in the KG (and to other KGs) can be explored via the "connective" power of the KG

  ○ Supports LLMs, AI, and Earth Digital Twin implementations

- NOAA is gearing up for promotion of OISS to the NCCF production tier. A cohort of about 10 staff are currently being trained on using Jupyter notebooks to configure the archive workflows.

- At NCEI they are also working on:

  ○ Pushing the OneStop-Inventory Manager to operational, which is sanctioned as the official NOAA Data Catalog, with metadata workflows refreshed across all NOAA Line Offices. CKAN will be decommissioned.

  ○ Cloud migrations, including science applications and data streams.

  ○ Moving the entire contents of the Storenext system (~ 5 PB) over the next 1.5 years, starting with teams in the Boulder, Colorado facility.

  ○ Initial focus on Passive Acoustic Data, Hydrographic Surveys, and Trackline Geophysics.

Discussion

- Doug Newman (NASA) noted that knowledge graphs first manifested as JSON-LD files. Has that been built? Ken noted this hasn't been done, as they need the public access piece first. It will take a little while for that to happen, as the internal team has to run some pilots and prototypes first. Ken will report back to WGISS once it is ready to share publicly.

- NASA's experience was that they built it and users didn't come. It will be interesting to see how NOAA goes.

- NOAA has a vast heterogeneous collection of datasets, and so through a properly trained LLM, users will hopefully be able to find what they need quicker. Will likely see it used initially within the digital twin work.

- Doug recognised the challenge is that this is a very new way to look at the problem - expanding from traditional search and discovery. Being able to spot connections not previously noticed has huge promise. Not yet sure what the use case is.

- Ken noted there are a couple of use cases NOAA is looking into, including a recommendation engine. When the knowledge graph is connected to the traditional search and discovery tools, some similar datasets the user might like can be suggested.

- Hoping to have a chat bot approach to search and discovery, with natural language based queries.

- Tom Sohre (USGS, WGISS Chair) noted the importance of semantics for knowledge graphs. Was NOAA able to take advantage of the previous WGISS semantics work?

- Distinct possibilities for NOAA work to feed into the interoperability work, especially around semantics.

|		
WGISS-57-05	Ken to ask his team about the relevance of the WGISS semantics in the NOAA Knowledge Graph work, or if there is NOAA semantics work that could be provided to WGISS to support the vocabulary (semantics) Interoperability Framework factor.	**Due: ASAP** **COMPLETE**

3.3 - ISO TC-211

Liping Di (ISO/TC-211) reported:

- ISO Technical Committee 211 (TC-211) is responsible for setting international standards on geographic information/geomatics. To date, they have published 96 standards, with 22 standards under development.

- Currently TC-211 has 39 participating member countries and 34 observing member countries.

- ISO TC-211 has set up liaison relationship with many related external organisations, including CEOS WGISS, which is a class-A liaison of ISO TC-211.

- One standard currently in development is ISO 19124: *Geographic information - Calibration and validation of remote sensing data and derived products*. Fundamentals (Part 1) has been published, with the SAR work (Part 2) in progress, and optical (part 3) and calibration and validation sites (part 4) to be started soon.

- ISO/TC-211 and OGC, together with CEOS, have formed the Analysis Ready Data (ARD) Standards Working Group (SWG). It is envisioned that the ARD standard will be a

multi-part ISO & OGC standard, using the CEOS Analysis Ready Data Framework as the basis.

− Part 1 covers the framework and fundamentals, and is expected to be published as a standard in June 2025.

− One difference between the CEOS-ARD specifications and the ISO/OGC standard is the use of 'sub-dataset metadata' rather than 'per-pixel metadata', as it won't be restricted to pixelated products.
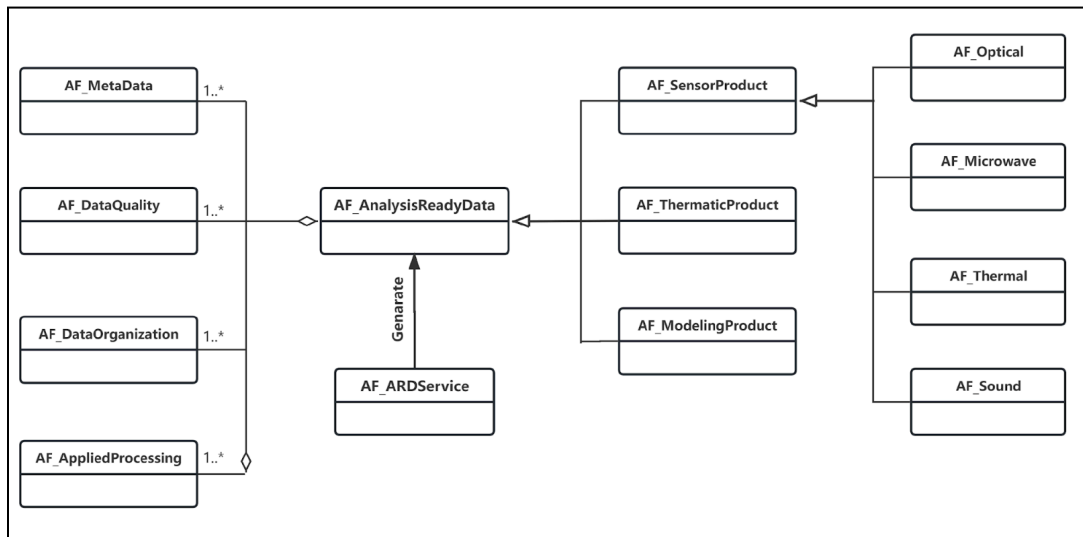
− WGISS members are welcome to join the SWG to develop the ARD standard.

Discussion

− Tom Sohre (USGS, WGISS Chair) asked about the future of the CEOS-ARD framework once the ISO standard is released. Liping responded that the current idea is that CEOS-ARD Oversight Group will restructure, and aim for maximal compatibility between the two. Will hopefully transition to the ISO standard.

− STAC is also being considered. The next OGC meeting is at the end of March, where STAC representatives will present on their work with the CEOS-ARD extension.

− Patrick Quinn (NASA) noted the ARD work is not very format specific at the moment, more about the content of the files. So far, the CEOS-ARD work hasn't been specific on the structure.

− Steve Covington (USGS) asked whether the CEOS-ARD concept of threshold and goal level would be included? Liping responded affirmatively.

– Simon Oliver (GA) recognised the shortcomings of the current CEOS-ARD specifications on specifying exactly how the measurements were produced, meaning there is incompatibility across different providers. Will there be more specificity on the measurements?

– Simon would like to see different products becoming more aligned. At the moment, it is open for people to choose their approach in reaching CEOS-ARD compliance.

– When Landsat Next launches, it will be incompatible with previous Landsat missions. Therefore, this equivalency issue should be addressed sooner rather than later, to improve the compatibility between Landsat Next and the Landsat archive.

WGISS-57-06	Simon to brief Liping and Patrick via email about the shortcomings with the CEOS-ARD product specifications, as input for the ISO/OGC ARD Standards Working Group.	**Due:** **7 March 2024** **COMPLETE**

3.4 - ISRO

Nitant Dube (ISRO, WGISS Vice-Chair) reported:

– INSAT-3DS launched on 17 February 2024, which is a replica of INSAT-3DR.
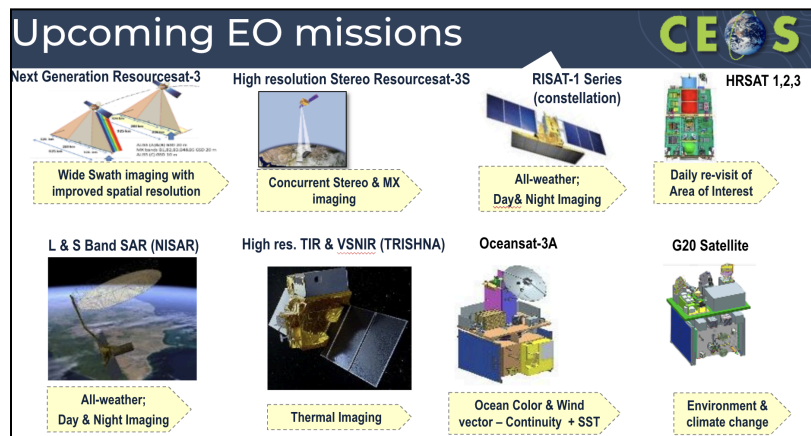
– Resourcesat is ISRO's land imaging satellite series, with RISAT the radar imaging series.

– The G20 satellite is specifically designed for environment and climate change issues, targeted at the G20 countries.

– ISRO hosts four data portals: Bhuvan, VEDAS, MOSDAC and Bhoonidhi.

– Bhoonidhi is a data repository of remote sensing satellites, facilitating the dissemination of satellite data products to online users on the web. Covers the data from 47 satellites over 33 years.

- The Indian Space Policy 2023 announced that all 5 m resolution and coarser data is available for free and open access to all users, available through Bhoonidhi. Finer resolution data is available for free to Indian government agencies, or commercial access otherwise.

- Spatial coverage depends on the satellite, some satellites cover the whole globe, while others are regionally focused.

- New Aerosol Optical Depth products over land, and oil spill trajectory forecasts are now available on MOSDAC. New INSAT-3DS data will be available on MOSDAC after in orbit testing is completed in the next couple of months.

- SPACE Data Visualisation and Analysis using iNteractive Computing Environment (ADVANCE) is a tool to promote open science. It provides Jupyterhub based services, and allows users to remotely login and process their data.

- ADVANCE uses Kubernetes clusters to support scalability. Jupyter notebooks can be published as web-apps using Mercury. It also allows interactive visualisation of data directly on the web using widgets. Python is used for carrying out processing and analysis of Space and Planetary data and downloading only the processed data/information.

Discussion

- Steve Covington (USGS) asked whether ADVANCE is open to the public? Nitant noted it will operate on the local cloud, with a plan to push onto the public cloud down the track. It will be integrated with the Bhoonidhi and MOSDAC platforms. All register users can use the platform - and anyone can register.

- ISRO replicates the Sentinel-2 data as well. Bhoonidhi is a Sentinel-2 regional hub, as well as for Landsat data acquired through their ground station.

- With ADVANCE, have users transitioned to using the data on the cloud, or do they prefer to download? ISRO internal users are comfortable with cloud processing, but there is a need to do some capacity development exercises for external users.

- With the opening of the ISRO archive, is computer storage becoming an issue? Nitant responded that this has not been raised as an issue yet, but it remains to be seen how it develops.

- Tom Maiersperger (USGS) noted that with the Digital Earth Australia and Africa sandboxes, they are open for anyone to use, with limited computing power. Users can apply for a grant to get access to more computing resources if needed.

- Matt Paget (CSIRO) noted that CSIRO has set up an infrastructure as a service system with a common structure to Digital Earth Australia, but a different business model.

3.5 - USGS

Steve Covington (USGS) reported:

- USGS has five key mission areas: Water resources, natural hazards, core science systems, ecosystems and energy and minerals. The Landsat programme falls under core science systems.

- Fundamental goal of the National Land Imaging (NLI) program is to ensure public availability of a primary data record about the historical condition and current state of the Earth's land surface and to predict its future condition.

- Landsat 1-9 focused on technical capability. In the second phase of the land imaging program, there was a wholesale rethinking of the user needs. From 2016-17, USGS did a comprehensive study of user needs to understand what observations are needed to support these applications.

- Landsat Next will have 26 bands, compared to the 11 bands on each of Landsat 8 & 9. This increase in bands was informed by user needs. Landsat Next will have a 6 day revisit, and spatial resolution up to 10 m.

- USGS makes use of domestic and international partnerships to enhance land imaging capabilities and increase quality and quantity of EO data available for users.

- Landsat 7 was to be used as a tech demo for the NASA OSAM-1 refuelling satellite, however on Friday the program was cancelled.

- There is a lot of overlap now of Landsat Next with Sentinel-2 and hopefully Sentinel-2 Next Generation as well. There is a need to have the same definition of what the different products entail, or the products won't be compatible.

- Needs to be better compatible between domestic US agency data systems. In designing USGS products, they try to think of the compatibility with other products. This requires working with the international community, including CEOS.

- Collection 3 will be released in the 2028 timeframe, and will include Landsat Next. Have about 24 months from now to define what the collection will contain. Want to discuss the product definitions with the international community to best accommodate interoperability.

- The current 50+ year USGS Landsat data is ~20PB. It is estimated that two years after Lansat Next launches, the archive will be close to 50PB. This is a big challenge both for providers and data users.

- Areas to explore to minimise risk:

    ○ Encourage processing and analysis "next to the data" (in the cloud) vs. data downloads

    ○ Enable streamlined multi-source, multi-modal data access capabilities

    ○ Understand opportunities (and challenges) for data replication; ensure quality and integrity of replicas

    ○ Focus on Earth Observation (EO) Data & Service Interoperability

- Last Friday, USGS and GA hosted a Landsat forum for local commercial users, who didn't seem to care too much about the accuracy.

3.6 - Discussion

- Should WGISS be creating Knowledge Hubs such as one like GA's? What would something look like this for all CEOS datasets?

- Steve Covington (USGS) suggested this could be something for the CEOS virtual constellations to tackle for their various domains.

- DOI provides the capability to have landing pages for the different products. Perhaps it is time to review the practices for the information provided in the DOIs.

- What kind of update would be needed to the DOI best practices? Could be ARD compliance, or ISO compliance. Quality indicators have been discussed at WGISS in the past.

- Steve would like to hear about NASA's perspective, on Ken's work and the implementation in general.

- Patrick Quinn (NASA) plans to reach out to Ken to connect some of the relevant NASA activities with those in NOAA.

- Doug Newman (NASA) noted NASA did go down the knowledge graph road, which looked promising, but wasn't overly successful.

- Data volumes will mean users will have to move to the cloud. Research can move much faster by using cloud based data.

− Steve recognised we don't have a good handle on the types of users we have. Only a few users take a lot of data - who are we designing our systems for?

− Lots of data is needed for training AI/ML models. Will see how we can create AI/ML ready data.

Session 4: Outreach and Communications

4.1 - External and Internal CEOS Collaborations

Tom Sohre (USGS, WGISS Chair) reported:

− The WGISS Terms of Reference define the key internal & external stakeholders:

> *Key external & internal stakeholders*
>
> The internal stakeholders of WGISS tend to be the data managers within the Space Agencies who are concerned with the preservation and the accessibility of data.
>
> WGISS also interacts with a broad range of stakeholders who are interested in seeing efficient data access and the provision of interoperable services. These include the other Working Groups, Virtual Constellations and other Ad-hoc Groups working within CEOS, together with external organisations, such as GEO, the Global Spatial Data Infrastructure Association (GSDI), the Consultative Committee for Space Data Systems (CCSDS), ICSU Committee for Data for Science and Technology (CODATA), commercial satellite data providers, and relevant bodies within the UN.

− Tom isn't sure whether this is relevant anymore, is it worth updating the Terms of Reference to reflect our recent connections?

− ISO, OGC and CGMS would be some other key people to refer to.

− Any update would have to go via CEOS Plenary for endorsement.

− CCSDS is still active, but GSDI is no longer active.

− The UN is still alive and well, but are they an external stakeholder for WGISS?

− STAC would be another key group, and ODC Steering Council.

− Recent internal CEOS Collaborations include:

  ○ SEO – ongoing collaboration with CEOS Analytics Lab and Open Data Cubes

  ○ LSI-VC – CARD4L, STAC

  ○ WGCapD – Jupyter notebook webinar

  ○ WGCV – joint meeting at WGISS-54 (and planned WGISS-58)

  ○ WGDisasters – joint meeting at WGISS-55

○ ARD OG – includes WGISS as a representative

− Should we have some better connections with these groups? Is annual reporting sufficient?

− WGISS used to do detailed work plans, and in the 2018 edition there was a table of organisations where WGISS should have a connection, and who the WGISS liaison is for the group.

− Mirko was the GEO POC in the past, and Makoto was the official WGISS representative. If there are relevant subgroups at GEO, we should consider specific POC for these as well.

WGISS-57-07	WGISS Chair to compile a table of POC for external and internal collaborations. Table should be regularly updated (revised at least annually) and may be included as part of a WGISS annual work plan.	Due: WGISS-58
WGISS-57-08	WGISS Chair to update the WGISS Terms of Reference (last updated 2015), specifically to address outdated Stakeholders section but other sections should also be reviewed as part of the update. An updated Terms of Reference will be presented at CEOS Plenary 2024.	Due: Q3 2024

− As part of WGISS-58, would it be worth inviting some commercial providers to have a discussion around these topics? Similar to what LSI-VC is doing in Tokyo next month. Would work with WGCV as well, and would try to get the providers to focus on specific topics.

− Matt Paget (CSIRO) agrees an industry day would be valuable, but it is important to understand what we want to get out of it. Is it WGISS responding to the NSTT white paper, or do we want to promote our standards and best practices, or understand their needs? We should work through these types of questions first.

− Could address both upstream and downstream sides - data providers and data producers.

WGISS-57-09	WGISS Chair to work with WGCV Chair to map out an outline for a Satellite Earth Observation Commercial Engagement Workshop at WGISS-58 / WGCV-54 in October 2024 modelled after the LSI-VC workshop to be held in April 2024 (using lessons learned from that workshop).	Due: WGISS-58

- Simon Oliver (GA) noted that a reason GA decided to try out the Landsat user forum last week was because they felt disconnected to the commercial sector.

- Nitant Dube (ISRO, WGISS Vice-Chair) noted that CGMS connected with cloud providers to develop the best practices documents. But this was focused on some specific topics.

- Yousuke Ikehata (JAXA) noted that there has been difficulties in finding AI ready data. But commercial providers may know the answer, as they are introducing AI/ML into their EO workflows. Commercial providers can share very powerful use cases. AI/ML white paper should make sure to include use cases from the commercial side.

4.2 - CEOS Communications overview

Libby Rose (SEO CEOS Communications Team) reported:

- The SEO supports CEOS communications, with support from a small group of agency contacts.

- In 2023, the new CEOS communications strategy was endorsed by CEOS Plenary.

- CEOS has a number of social media channels where CEOS work is promoted. This includes Twitter, Facebook, LinkedIn and YouTube.

- The CEOS Blog can be found at ceos.org/news, and provides a space to share work from across CEOS, as well as highlight thematic topics. Published articles also appear on the front page of the CEOS website. This is a great way to give visibility to WGISS work.

- The quarterly revisit was started last year, and is a quarterly email newsletter summarising the recent articles and other topics such as training.

- The strategy outlines the key target audiences for CEOS communications. Broadly, it is categorised into two buckets: Internal and External.

- For Internal, this includes CEOS Principals, aiming to show them the value for money that WGISS brings, and working teams, ensuring cross cutting work like the interoperability framework is shared broadly between the groups.

- On the external side, WGISS is probably likely to target the EO Experts group most, including data providers, commercial industry and scientists.

- Over 2024, a number of campaigns will be run - which are a series of connected materials covering a common theme.

- The first is about the 40th anniversary - 2024 marks 40 years since the first CEOS Plenary. WGISS has been a key aspect of CEOS throughout the whole 40 years, with the

Working Group on Data, as WGISS was formerly known, one of the two original working groups, alongside WGCV.

- The Biodiversity campaign focuses on supporting the CEOS Chair's priority, while the GHG campaign supports the SIT Chair priority.

- The SEO also supports CEOS booths at various events, including IGARSS and GEO Week. There is no GEO Week this year, but Libby will attend IGARSS in July in Athens, alongside Dave Borges. This is a really good way to connect with the community, and WGISS could think about ways to promote their work, through a flyer or otherwise.

4.3 - Discussion: WGISS Outreach and Engagement Goals

- The following questions were raised for discussion:
  - What are the goals of WGISS for outreach?
    - How can we support this with outreach and communications?
  - What does WGISS want to promote to the broader EO community?
    - What one thing should be put on a flyer?
    - Are there different topics which should be targeted towards data providers vs data users?
  - How can we bring new people in to support WGISS work?
  - Do we need a generic "What is WGISS?" slide deck?

Discussion

- Tom Sohre (USGS, WGISS Chair) agreed that a generic slide deck would be useful. Target audience would probably be management at different organisations.

- Doug Newman (NASA) asked what are the success stories of WGISS? CWIC could be an example.

- Nitant Dube (ISRO, WGISS Vice-Chair) suggested a video for WGISS activities. Yves Coene (ESA/Spacebel) noted the video made for IDN.

- The goal of WGISS outreach is for the promotion of activities within management chains at agencies.

WGISS-57-10	Libby to work with Exec on developing a standard WGISS overview presentation. Each IG Chair should provide at least one slide describing activities within	**Due: WGISS-58**

	the IG. The Chair Team will provide slide(s) describing WGISS overall goals and recent accomplishments.	
WGISS-57-11	Libby to update the brochure link online to the more recent version provided by Iolanda. Consider whether to remove the brochure once the slide deck in WGISS-57-10 is complete.	**Due: ASAP**

Session 1: Opening Session (cont.)

1.5 - CEOS Executive Officer Report

Steven Ramage (CEOS Executive Officer) reported:

−   Steven has been involved in the geospatial industry for many years, most recently with GEO and OGC. He started in the role of CEOS Executive Officer in January this year.

−   Steven is working with Evenflow as support for the role. Lefteris Mamais and Irena Drakopoulou will be the primary support from Evenflow.

−   The CEOS Executive Officer:

    ○   Develops the CEOS Work Plan

    ○   Discusses CEOS contributions to the GEO Work Programme (supports SIT Chair on GEO Programme Board and sat on Executive Committee)

    ○   Advises CEOS leadership on continued/increased internal and external cooperation

    ○   Tracks progress on the CEOS Work Plan Objectives and Deliverables

−   The focus is on preparing for SIT-39 next month, and also the GEO Programme Board next week.

−   GEO has put a call out for members for an AI Subgroup.

−   The CEO is supporting internal communications and outreach, notably around the 40th anniversary of CEOS.

−   The CEO has been working with WGCapD on engaging more emerging economies (outside of Europe and North America), as well as with WGDisasters on the international Early Warning for All programme.

−   Externally, Steven will also join the OGC Board of Directors later this month on behalf of CEOS. There are numerous topics in OGC around ARD, STAC, New Space, Interoperability, OpenEO etc, which are all relevant to WGISS.

- The Taskforce on Nature-Related Financial Disclosures (TNFD) approached the CEO to discuss biodiversity monitoring from space

- The CEO is working externally with the World Economic Forum (WEF) on how to value Earth observations. GEO is also working on an EO Valuation Toolkit.

- The CEO has also been approached around disaster risk and more use of SAR in the International Disaster Charter by the UN World Food Programme (WFP), as well as for process flow improvements by the Humanitarian OpenStreetMap Team (HOT).

- WGISS is in great position with the Work Plan, with the required sections updated.

- Steven is trying to create a more integrated approach for the Work Plan.

- The CEO is attending the following events in the first half of this year:

  ○ SIT-39, Tokyo (April 2024)

  ○ GeoIgnite, Ottawa (May 2024)

  ○ SatSummit, Washington DC (May 2024)

  ○ World Biodiversity Summit, Davos (June 2024)

  ○ UN COPUOS, Vienna (June 2024)

  ○ IGARSS, Athens (July 2024)

Day 2: Tuesday 5th March, 2024

Tom Sohre (USGS, WGISS Chair) reviewed the actions from Day 1. One note regarding the outreach session is the target audience for WGISS activities. If the audience is just CEOS Agencies internally, that is okay, but we should be clear on what this is, especially when it comes to the data access tools.

Richard Moreno (CNES) suggested that the target audience should be anyone hosting their own data. Not necessarily limited to CEOS Agencies, but also new space.

Steve Covington (USGS) suggested WGISS should think about finding ways to unify approaches across data providers to improve exploitation of like data sets. Audience is the agencies, not the users. WGISS members should represent their respective user communities.

Session 5: Interoperability

5.1 - Work Plan for Data Interoperability and Use Interest Group (DIIG)

Nitant Dube (ISRO, WGISS Vice-Chair) reported:

− In 2023, the Interoperability Framework and Roadmap was completed. A second version of the Interoperability Handbook (originally produced in 2008) will be produced in 2024. Emails have been circulated to all CEOS group leads to identify contributors for the handbook.

− In 2025, the group plans to work on a maturity matrix for interoperability.

− Possible contributors from WGISS and other groups have been identified:

Factor	Vocabulary (Semantics)	Architecture	Interface (Accessibility)	Quality	Policy
WGISS	TEIG, DIIG	DSIG, DIIG	**DAIG**, DIIG	DIIG	DIIG
Other WG	WGCapD, WGCV, CEOS-ARD, LSI-VC	WGCapD, WGCV, CEOS-ARD, LSI-VC, SEO	CEOS-ARD, LSI-VC	CEOS-COAST WGCV, NSTT, **GEO (QA4EO)**	NSTT, **CEOS/Agencies**

− The last two factors aren't clear with who from WGISS should contribute. Contributors are welcomed.

− Thanked Prasanjit Dash from NOAA and representing CEOS-COAST for nominating to contribute to the quality factor.

Work Plan 2024-2026

Year	2024	2025	2026
Interoperability Framework	Handbook V2.0	Handbook V2.0 + Maturity Matrix	**Interoperability Demonstration** : Development of Earth Observation Plug and Play (EOPnP) modules

Year-2024	Q1	Q2	Q3	Q4
Handbook V2.0	E-mails sent, with request for identification of Champions	Discussions with Champions/ leads and Draft Sketch of HB 2.0	Detailed discussions on each Factors	Presentation of Draft HB2.0 during WGISS-58

Year-2025	Q1	Q2	Q3	Q4
Handbook V2.0 + Maturity Matrix	Presentation of Final version of HB2.0 during WGISS-59 Maturity Matrix Discussions during WGISS-59	Release of HB 2.0 for comments Development of Maturity Matrix	Incorporation of Comments Development of Maturity Matrix	Endorsement during CEOS Plenary Presentation of Maturity Matrix during WGISS-60

- In 2026, DIIG would like to demonstrate interoperability via an Earth Observation Plug and Play (EOPnP) module. This would hopefully be done in collaboration with the CEOS Analytics Lab, although this is yet to be defined.

WGISS-57-15	DIIG to contact Brian Terry & the SEO team regarding the use of the CEOS Analytics Lab for the EO Plug and Play (PnP) demonstrators.	**Due: WGISS-58**

Discussion

- Mirko Albani (ESA) asked about the Maturity Matrix, and the distinction between the Data Management and Stewardship Maturity Matrix (DMSMM) and the proposed Interoperability Maturity Matrix. Tom Sohre (USGS, WGISS Chair) noted the intent would be different, but it is still conceptual. The goal is to create something to help organisations to understand where they sit on the spectrum of interoperability.

- Should make sure we are clear on the wording used to not be confused with DMSMM.

- The interoperability matrix will likely refer to DMSMM for the data access piece.

- Steve Covington (USGS) noted some concern with the timeline. The first real opportunity to apply these concepts will be for Landsat Collection 3 which is being defined at the moment. Want to contain the large-scale aspects of this interoperability work in Collection 3, but it is unclear how the timing will work.

- Nitant noted they want to do away with static documents. Interoperability is such a big job, won't be able to do a complete job for 2.0. Once collection 3 lessons have been applied, can do a 2.1, 2.2, etc.

- Tom recognised that learning from demonstrators will be key to develop the handbook and maturity matrix.

- Alex Leith (SEO/Auspatious) noted that for a lot of the modern standards, rather than starting by writing the specifications, they start with a demonstration and then building the specifications from there. Perhaps the 'Plug and Play' modules should be brought to the start and done alongside the handbook development.

- Damiano Guerrucci (ESA) is interested in the plan, and specifically the Plug and Play principal. Are we looking at defining some modules at a CEOS level for further implementation, or stopping at a demonstrator level?

- If the Plug and Play principals work well, other Working Groups or Agencies could continue building on the model.

- Yousuke Ikehata (JAXA) asked about the TEIG contributions. TEIG was listed under the vocabulary box, but this hasn't been discussed yet within their group.

- In the next meeting of DIIG, the team will start getting into the details, and working out who will contribute where. IG leads can contribute to multiple factors, and should contribute across the board where necessary.

- TEIG could contribute around thesaurus or knowledge graph work.

5.2 - Interoperability Framework and Roadmap

Tom Sohre (USGS, WGISS Chair) reported:

- At the 2022 CEOS Plenary, CEOS Principals confirmed that coordination of interoperability related work remains within the WGISS perimeter. WGISS was invited to propose an interoperability roadmap at SIT-38, considering the CEOS Interoperability Framework concept developed and presented by LSI-VC. (Decision 36-09).

- WGISS coordinated across CEOS groups (ARD Oversight Group, LSI-VC, SEO, WGISS, WGCV, etc.) to develop a framework and mapping of CEOS activities / groups / people to each interoperability factor. This was presented and endorsed at CEOS Plenary 2023.

- It wasn't until WGISS-55 that it was recognised that WGISS had previously developed an Interoperability Handbook in 2008, however it was out of date.

Proposed Interoperability Framework

Vocabulary (Semantics)	The (narrow) semantic aspect refers to the naming and meaning of data elements. It includes developing, harmonising, and maintaining vocabularies and schemata supporting provision, exchange, and analysis of data, and ensures that terms and data elements are understood in the same way by all communicating parties
Architecture	Architecture describes the organisational structure of concepts, processes, and assets, including data and workflows. It comprises the structural aspects of models and standards that govern the collection, storage, arrangement, integration, and use of data.
Interface (Accessibility)	Data exchange protocols, and application interfaces. These provide the means necessary to access and exchange data.
Quality	References are data and schemes that are used as benchmarks for (observational) data comparison or analysis. This could include instances such as geographic locations, product numbers, or official (authoritative) data and statistics.
Policy	Legal frameworks, policies and strategies regulating the relation between the different stakeholders.

− Note that these are not black-and-white boxes, there is a lot of overlap between the factors different activities support.

− The team also looked across the CEOS Work Plan to identify activities which support each of the five factors.

− Can't look at it from 'solving' interoperability, only from the aspect of improving interoperability.
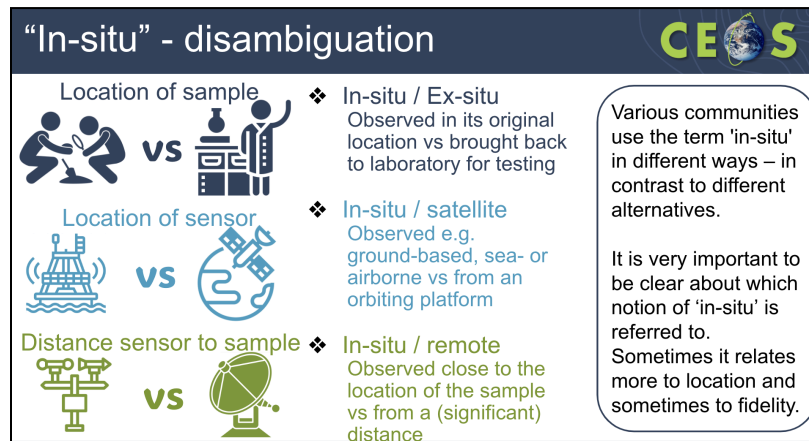
Discussion

− Simon Oliver (GA) asked where measurement equivalence would fit in, as it is not necessarily a quality issue. Perhaps this could fit within architecture? Ability for the measurements to work together.

− A lot of the activities probably have a secondary or tertiary alignment within the framework.

5.3 - Vocabulary (Semantics) - CEOS Common Dictionary

Katrin Molch (DLR) reported:

− Inconsistent use of terminology by different communities hampers communication in interdisciplinary research and other collaborations.

− Project-specific glossaries with altered/new definitions to overcome the issue often make things worse.

– CEOS WGCV recorded an action item in 2021 to advance the idea of a CEOS common online dictionary, with a view to eventually reaching out to launch a broader community effort.

– The work is considered completed, with a publication submitted to Surveys in Geophysics in December 2023 to report on the activity and to summarise the results. The publication is currently in peer review, with preprint available upon request. The paper is titled *"Lost in translation: The need for common vocabularies and an interoperable thesaurus in Earth sciences"*.

– The group reviewed 12 existing dictionaries, and considered their useability, structure and identification of circular definitions and inconsistencies.

– Reviewed ISO 25964:2011 guidelines on 'information and documentation: thesauri and interoperability with other vocabularies'

– Studied in detail selected sample terms - observation, in-situ, interoperability - as examples of broader problems.

– A hierarchical set of base and core terms was built.

– Wide interest in the topic was noted wherever it was discussed. The need and urgency are clearly recognized as is the huge dimension of the undertaking – if it is to be done well.

– Many individual, well-designed dictionaries exist, but they mostly exist in isolation; not linked, with no overarching effort or framework. This isolation creates inconsistent, controversial and superficial definitions.

– Criteria were proposed for a good dictionary: it should be consistent, hierarchical, understandable, educational and updateable.

– ISO 25964:2011 provides an excellent framework for building a consistent dictionary

- A Wiki-like system, moderated by expert groups, might be the most suitable implementation. One unique source, accessible, easy to reference, editable, and moderated

Possible way forward

❖ Start discussions about formulating a properly funded project to establish an example of a public dictionary
    □ CEOS with its Interoperability Framework could be a suitable body to coordinate such an effort □ Semantics is a key factor in the framework

❖ Update the CEOS Cal/Val portal definitions with
  ▪ Hierarchical framework for the 30 core terms defined so far as an example
  ▪ Highlighting and linking core and base terms within

❖ Connect to CEOS interoperability framework

CEOS Interoperability Factors

Semantics	Architecture	Interfacing	Quality	Policy
Naming and meaning of terms and data elements, includes developing, harmonising, and maintaining vocabularies and schemata	Organisational structure of concepts, processes, and assets, including data and workflows	Data exchange protocols, and application interfaces necessary to access and exchange data and workflows	Performance against references which are data and schemes that are used as benchmarks for comparison or analysis	Legal frameworks, policies and strategies regulating the relation between the different stakeholders

Illustration courtesy P. Strobl

Discussion

- Steve Covington (USGS) recognised that the only way to create a real long term consistent lexicon is to make sure everyone is trained on that lexicon. Unless we have a clear communication strategy for the dictionary, it won't be successful.

- Tom Sohre (USGS, WGISS Chair) noted the paper will be a key component of the new handbook. Could a large portion of this paper be adopted for the semantics chapter of the handbook?

WGISS-57-13	Interoperability IG Lead to work with Katrin to incorporate components of Common Online Dictionary Research Paper into the Interoperability Handbook 2.0 Vocabulary Section.	**Due: WGISS-58**

- Nitant Dube (ISRO, WGISS Vice-Chair) would like to put some key recommendations into each of the chapters, which is where the paper could be useful.

- Tom asked whether there is an understanding of how many dictionaries are just within CEOS? The DSIG document is a key one, but not very user friendly as it is just a PDF document. It might be an unsolvable problem to get everyone in the world to agree, but if we can get the CEOS documents to be synchronised, it would be a good first step.

- Katrin noted that scope of the discussion was a key part of the work, and it was limited to a 'CEOS Common Dictionary'. Within CEOS, we should have something that encompasses all.

5.4 - Geospatial Standards and their role in Interoperability

Liping Di (ISO/TC-211) reported:

- Geospatial resources include hardware, software (process), and geospatial data and information.

- Interoperability is the ability of two or more systems from multiple vendors to communicate (i.e., exchange information and/or services) with each other meaningfully.

- Geospatial interoperability means that geospatial systems from different vendors can communicate and collaborate to fulfil a geospatial analysis task.

- In order to communicate meaningfully, individual parties involved should agree on the common methods for communication among them. The common methods that everybody agrees upon are the standards.Therefore, standards are key for geospatial interoperability.

- Standards are statements of fact, quality, procedures, or content, to which applicable entities are compared for purposes of acceptance and/or use. They are documented agreements that contain or specify technical or other specific criteria to ensure that processes, products, or services meet their intended purpose.

- Standards can be designated as either voluntary or mandatory with respect to usage.

- Using common standards increases the shareability, reliability, and effectiveness of geospatial data.

- Geospatial standards can be categorised by the following:

  - Data Classification: provide groups or categories of data that serve an application.

  - Data Content: provide semantic definitions of a set of objects.

  - Data Symbology or Presentation: define graphic symbols.

  - Data Transfer: independent of technology and applications and facilitate moving data among systems, without prior specification of the intended end use of the data.

  - Data Useability: describe how to express the applicability or essence of a data set or data element and include data quality, assessment, accuracy, and reporting or documentation standards.

– Stakeholder participation in the standard development process is key for making standards useful in the interoperability.

Discussion

– Nitant Dube (ISRO, WGISS Vice-Chair) noted that in addition to the data and services, it is interesting that knowledge interoperability is a key component for ISO/TC-211.

5.5 - SEO Report

Alex Leith (SEO/Auspatious) reported on behalf of Dave Borges (NASA, SEO):

– The CEOS Analytics Lab (CAL) is a Jupyter lab deployment, for collaborative work at a technical level.

– Anyone who wants access to use CAL should reach out to Dave and the SEO team. CAL can be found at ceos.org/cal.

– One of the key aspects for CAL is around interoperability, and testing out interoperability for various new space datasets.

– New space datasets are currently being ingested into CAL, including both SAR and optical datasets.

– The technical approach to this work includes:

  ○ Evaluating the use of commercial provider APIs (not all include API access), and implement APIs to access data

  ○ Generating ODC indexing scripts for each commercial provider

  ○ Creating demonstrator indexer notebooks

  ○ Evaluating the pixel-by-pixel scatter plots comparing individual bands, Inherent harmonisation, resampling analysis, and band to band spectral comparison.

– Entire Landsat and Sentinel-2 archives are currently available through CAL.

– On the CAL website, there are service and data request links.

– Another tool in the CEOS suite is the CEOS Visualisation Environment (COVE) (ceos-cove.org). There are a number of tools available within this environment.

– The SEO is keen to learn about the use of COVE within WGiSS.

– CAL is not open, one needs to request an account. Primarily for use within the CEOS community.

– Dave will be attending SatSummit in May and IGARSS in July for outreach.

Discussion

− Simon Oliver (GA) noted he finds COVE very useful, and would like to see a spectral response function.

− Are there some metrics on usage? This would likely be a better indicator of its usefulness.

WGISS-57-14	Libby to follow up with SEO regarding metrics for COVE use and access to WGISS Best Practices and Guides.	Due: ASAP

− Filippo Marchesi (ESA/Solenix) noted that at the DAIG meeting, they tried to go into the COVE tool, but access was required. Is there a way of federating access?

− Matt Paget (CSIRO) many of these tools would have openID as a login option. But it can be tricky to manage the user authorisation. Perhaps there could be a central CEOS authentication aspect.

WGISS-57-15	DAIG to discuss with SEO the possibility of a central CEOS Authentication mechanism.	Due: WGISS-58

− A notebook to illustrate the use of the API could also be helpful.

− Damiano Guerrucci (ESA) noted COVE wasn't known very well to WGISS, and would like to understand the tool a bit more. It would be good to look at data interoperability within COVE, and with CAL.

5.6 - CEOS-ARD: New Strategy for 2024

Matt Steventon (LSI-VC Secretariat) reported:

− Matt is reporting on behalf of the ARD Oversight Group lead, Ferran Gascon (ESA), and the LSI-VC co-leads, Steve Labahn (USGS), Andreia Seiquira (GA) and Peter Strobl (EC-JRC).

− CEOS-ARD datasets are satellite data that have been processed to a minimum set of requirements and organised into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets.

− The motivation is to ease the use of satellite EO data to reach new non-expert users and increase data uptake and use.

- Five product family specifications currently exist: Surface Reflectance, Surface Temperature, Aquatic Reflectance, Nighttime Lights Surface Radiance and the combined SAR specification.

- There is interest from other VCs to expand the list of products to beyond land, including precipitation and ocean colour. Interferometric SAR and LiDAR PFS are also in development.

- The CEOS-ARD SAR team has combined the existing PFS for CEOS-ARD SAR products into one document to ensure consistent parameter names and specifications across all SAR PFS. This includes normalised radar backscatter, polarimetric radar, ocean radar backscatter and geocoded single look complex.

- The CEOS-ARD Strategy 2024 is in draft at the moment. It is an update from the 2019 and 2021 versions of the strategy.

- Some aspects of the strategy directly related to WGISS include:

  - Maximising the discoverability and accessibility of CEOS-ARD by placing CEOS-ARD in the cloud, easily identifiable, discoverable, accessible, and in close proximity to users and their tools, and ensuring maximum 'machine-to-machine' functionality.

  - Advancing interoperability of CEOS-ARD by leveraging the CEOS Interoperability Framework and its principles.

- The strategy has five broad themes. WGISS support will be key for the framework and specification advancement chapter, and the discoverability, access, utilisation and interoperability chapter.

- Under chapter 2: Framework and specification advancement, the sections are:

  - 2.1: Evolution of the CEOS-ARD Framework

  - 2.1.2: CEOS-ARD product Levels

  - 2.1.3: Terminology

  - 2.1.5: Product Family Specification and CEOS-ARD Product Version Control

  - 2.2: Metadata specifications for optical specifications

- Under chapter 3: Discovery, access utilisation and interoperability, the sections are:

  - 3.1: Embrace the CEOS Interoperability Framework

  - 3.2: Adoption of cloud native technologies and facilitate machine-to-machine access and utilisation

- ○ 3.4 CEOS-ARD discoverability and branding
- ○ 3.5: CEOS-ARD on the cloud
- ○ 3.6: Technical Advisory Notes

- − The CEOS-ARD Oversight Group would be looking to work closely with WGISS to embrace the interoperability framework.
- − Want to make sure the work between CEOS-ARD and WGISS is coordinated well. Inputs are welcomed on the draft CEOS-ARD Strategy 2024.

Discussion

- − Tom Sohre (USGS, WGISS Chair) asked how the draft will be progressed. Matt noted that it will be developed by the ARD Oversight Group alongside the Virtual Constellations.
- − There are lots of good actions within each of those sections, and it might be good to help the coordination to understand who will lead each of the items, including if there is an expectation for WGISS to take one of those items and lead.

WGISS-57-16	ARD Oversight Group to identify where specifically WGISS should support the activities identified in the 2024 CEOS-ARD strategy.	Due: Q2 2024

- − Steve Covington (USGS) asked about the strategy for engaging with the commercial industry.
- − There is a New Space engagement workshop at LSI-VC-15 in April, but Matt hadn't thought about presenting the strategy to them as well, as he saw it more as internally to CEOS. It could be useful to get their opinions.
- − A chapter 6 on commercial engagement was suggested, as an evolution of the current CEOS-ARD industry engagement strategy.
- − AIR data is a good test case to see whether the specifications are a good test to see if the metadata standards are sufficient to discriminate the AIR Landsat product from the USGS Landsat product. Having metadata that specifies the chain of traceability should be included in the framework.
- − Nitant Dube (ISRO, WGISS Vice-Chair) recognised that ARD will be a good demonstrator for interoperability. Is anyone from the ARD Oversight Group ready to contribute to the interoperability handbook? Specifically on the architecture factor, which was promoted specifically from the ARD side.

– Matt confirmed that LSI-VC and the ARD Oversight Group can be counted on to contribute as needed.

WGISS-57-17	Interoperability IG Lead to set up a meeting with LSI-VC and ARD-Oversight Group leads regarding someone from those groups taking a leadership role for development of the Architecture component of the Interoperability Handbook 2.0.	**Due: Q2 2024**

5.7 - Cloud Native Geospatial and Interoperability

Alex Leith (SEO/Auspatious) reported:

– Digital Earth Australia formed out of the Open Data Cube project, which was in turn built off the Unlocking the Landsat Archive project. It aimed to lower the barrier to entry to use EO data.

– Alex also worked on the Digital Earth Africa project, with the broad goals around making data more easily accessible across the African continent. It leant on the readily available Level 2 products from Sentinel-2 and Landsat Collection 2.

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	✗	✗	✅

– Global ARD products are available with low to no barriers to entry.

– There are petabytes of open data available, alongside new commercial satellites almost every week.

– The standard STAC API allows for an increased interoperability for search and discovery.

- A STAC Catalog contains a list of catalogs or collections. A STAC Collection contains items, which is a single space/time/thing. Items contain Assets, which is a link to a resource.

- STAC API is a REST API that contains a searchable interface for collections and items.

- There is a big demand to be able to easily access EO data on the cloud. Alex demonstrated that with 6 lines of code and 2 python libraries, satellite data can be loaded from anywhere in the world.

- For Digital Earth Pacific, Sentinel-1 mosaics were built. Sentinel-1 data over a year has about 60 ascending and descending scenes. To compute a mean, all the data must be collected. To do it more efficiently, can download all the data, process it just once and store the values locally.

- Picked up the Digital Earth Australia and Africa coastline workflows to other regions of interest. Did the work for Vietnam on the fly, loading Landsat scenes via the USGS STAC API, the Copernicus DEM via MSPC STAC API, and the ESA WorldCover product.

- Process of using the STAC API means the process can be applied anywhere in the world.
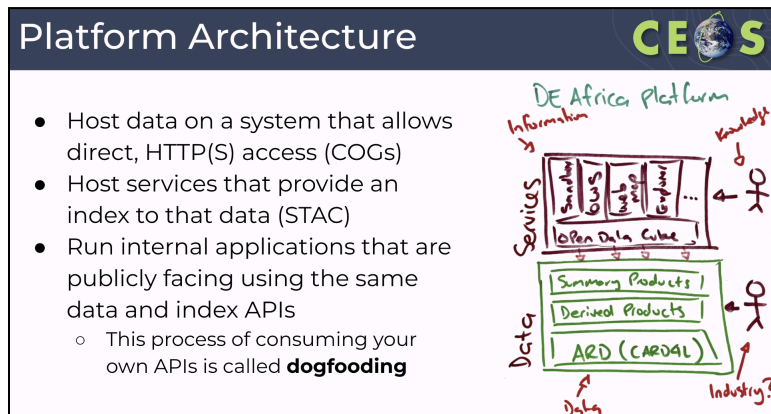
- If you have a STAC API and it is used for internal workflows, one can be confident it can be applied externally.

- Cloud Native Geospatial Foundation encourages adoption of highly efficient and accessible approaches to working with geospatial data over the Internet.

- The SEO is hosting a workshop at SatSummit in mid-May 2024, focusing on accessing cloud native geospatial datasets.

WGISS-57-18	SEO (Alex Leith) to share information with WGISS about the SatSummit workshop coming up in mid-May 2024.	Due: ASAP

Discussion

− Tom Maiersperger (USGS) recognised there are still some barriers to entry, including the need to build an environment to use and access the data.

− Alex challenged this, noting that ODC was fundamental to enabling the access. Users can access the data with only python and a couple of libraries, and no AWS account is needed.

− Tom Sohre (USGS, WGISS Chair) recognised that the missing piece is the education aspect about just how easy it can be to access data.

WGISS-57-19	WGISS Exec to discuss an approach to sharing data access success stories (such as Alex's code snippet demonstrating cloud native data access) and how we could facilitate instructions to stakeholders.	**Due: Q3 2024**

− Steve Covington (USGS) noted agencies need to work to change the behaviour of users in the next six years, ahead of the Landsat Next launch.

− Matt Paget (CSIRO) recognised the ease of access is only possible because the Landsat archive is COG enabled and placed in a STAC API.

− George Dyke (SEO/Symbios) noted that commercial providers are not necessarily looking for easy access to their data because of the commercial nature.

− Steve Covington (USGS) asked whether WGISS can work with cloud vendors to limit egress fees? Cloud vendors would make money off the CPU use instead.

Chat Messages

Alex Cesarz (CDSE): use (computing) of EO data is very limited and does not justify the storage costs. Keeping >100PB of data and necessary infrastructure to manage and serve this data needs sustained funding if we want to get rid of egress cost.

5.8 - Discussion: Knowledge Graphs

Nitant Dube (ISRO) reported:

− The migration from Web (2.0) to Semantic Web (3.0) makes data more machine readable.

− Linked data allows for the sharing and linking of machine readable data on the web (http/https). It includes a Uniform Resource Identifier (URI), Resource Description Framework (RDF), stored in a graph database.

- Ontology: Generalised data Model for defining objects and their relationships; RDF is used to define ontology.

- Graph of interconnected real world objects created using ontologies is called a Knowledge Graph.
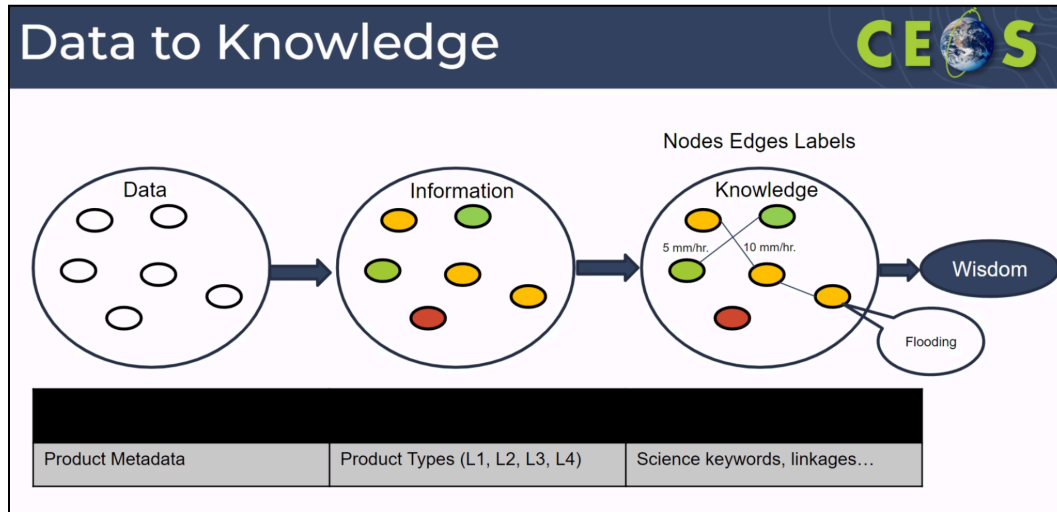
- Applications include data recommender systems, knowledge graph based question-answering systems, natural language based metadata search, AI/ML, and ultimately is about generating deeper insights about data.

- Data discovery and access from heterogeneous sources will improve.

Yves Coene (ESA/Spacebel) reported:

- FedEO is part of WGISS Connected Data Assets. There are a couple of research projects related to Knowledge Graphs (KG) in the context of FedEO.

- EOVOC and E-MC projects supported by BELSPO are exploring RDF Graphs and GraphQL. They aim to provide a Semantic Layer/Knowledge Graph on top of FedEO metadata layer.

- RDF Knowledge Graphs is part of a linked open data cloud, with all resources identified by URI. They use both the GeoDCAT-AP vocabularies/properties and the Schema.org vocabulary.

- FedEO "feeds" search engine Knowledge Graphs via structured data.
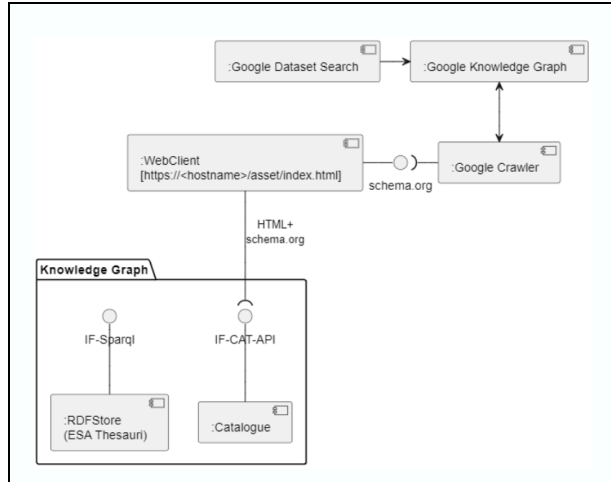
- – GraphQL originated at Facebook in 2012.

- – The Schema Definition Language defines request and response structures. Users can just ask for specific properties of interest.

- Implementations of these knowledge graphs require further work. RDF/SPARQL and GraphQL are not sufficient to achieve interoperability. Same RDF vocabularies and SDL schemas need to be used.

Discussion

- GraphQL is a browser tool, and outputs an array of data possibilities.

- Patrick Quinn (NASA) noted that NASA uses this extensively. When a specific term is queried, the knowledge graph can give similar terms and the number of matches. Can construct SQL-like queries, and the server then decides how to fulfil the request.

- STAC API is going in a similar way, but a different direction. There is some sort of meshing of the two which ought to happen.

- Good for less technical people to query data, including people who are developing tools.

- Doug Newman (NASA) recognised WGISS needs to understand knowledge graphs before trying to work out how to make them interoperable. People aren't yet understanding the capabilities.

- Steve Covington (USGS) asked whether there is something USGS should build on top of Landsat data? Doug noted probably, but it isn't yet clear how.

- Use case is to extract new knowledge from existing knowledge. Doug was hoping to take NASA's datasets, and link it out to abstracts to find connections between collections. His team found a lot of shared documentation between the collections.

- If WGISS will look at this any further, perhaps it is a technology exploration area. Not yet sure it is at the level of interoperability framework or data access.

- It was suggested to connect with the MIM Database as well to make connections across the different agency datasets.

- Doug noted that NASA made the choice not to go down to the file level, as it was seen as too costly. The community wasn't leveraging the tool to the extent to make it worthwhile.

- Richard mentioned a couple of use cases from Europe which tackled these topics.

Session 6: Technology Exploration

6.1 - Work Plan for Technology Exploration Interest Group (TEIG)

Yousake Ikehata (JAXA, TEIG Lead) reported:

- The 2023-2025 work plan included the Jupyter notebook best practices and the AI/ML work plan.

- The updates made this year include updating the delivery date for the jupyter notebook best practices to 2024, and adding federation topics, including knowledge graphs.

Number	Deliverable	Projected Completion
DATA-22-01	Jupyter Notebook Best Practice	2023 Q4 **-> 2024 Q2**
DATA-23-01	AI/ML White Paper	2024 Q4

- There are a few comments to address with the jupyter notebook best practices ahead of the deadline in Q2 2024. TEIG needs some more support from WGISS broadly to resolve these comments.

- Yousuke will upload this to GitHub to help the WGISS community address these comments.

- The AI/ML white paper introduces activities for analysis of Earth observation data using computational and data science methods, such as machine learning (ML) and artificial intelligence (AI) that various CEOS agencies have conducted to date.

- The outline has been written and some contributors identified. However, they have been unable to identify someone from the European Commission to contribute to the AI/ML white paper.

Discussion

- Steve Covington (USGS) recognised that the Jupyter Notebooks best practices is a good first step, but what comes next? Not sure if the target is toward data providers or data users. How can users best access the datasets, and how can the data providers best format the data for use?

- Tom Sohre (USGS, WGISS Chair) noted the best practices focus on the lessons learned from two previous webinars done in coordination with WGCapD.

- Steve added that USGS want to head towards a change in data use, with a move towards using the data in the cloud. Would hope the next step would be to plan out a strategy to do so.

- Alex Leith (SEO/Auspatious) asked how the best practices fit in with capacity development? Have you gotten feedback from notebook repositories, for example, Digital Earth Australia?

- Nitant Dube (ISRO, WGISS Vice-Chair) noted the best practices describe to a data provider the practices to follow when writing a notebook. If compliant, should we make a statement within the repo? But who would do the certification? Perhaps it could be a self-compliance process.

- Tom Sohre (USGS, WGISS Chair) asked whether we can make some code snippets and broader outreach for the jupyter notebooks? Demonstrate the benefits of these best practices.

WGISS-57-20	Libby to work with TEIG on some communication materials around the Jupyter Notebook best practices.	**Due: Q2 2024**

- Nitant suggested a demonstration of what the best practices bring to users, and how it is beneficial.

WGISS-57-21	TEIG to contact EUMETSAT for some example notebooks to demonstrate the Jupyter Notebook best practices. Look at WGISS-56 presentations and FedEO catalogue.	**Due: ASAP**
WGISS-57-22	TEIG to work with SEO to ensure CEOS Analytics Lab notebooks are compliant with the Jupyter Notebooks best practices.	**Due: Q2 2024**

- Tom asked about the next steps for the AI/ML White Paper.

- Yousuke will plan to publish the white paper as a first version, and look to involve the community for comments, including inputs from other WGISS activities. White paper will be updated periodically.

- Would like the various agencies doing AI/ML work to review the paper and provide feedback.

- Alex Leith (SEO/Auspatious) noted one of the Digital Earth Africa products was a crop probability product. This uses random forests and AI/ML technology and could make for a good complete use case. Radiant Earth has been collecting training data for quite a long time.

- Steve Covington (USGS) asked whether WGISS should reach out across CEOS to ask about AI/ML applications, including perhaps WGClimate?

- The purpose of the white paper was to document the use of AI/ML in EO applications, and some of the challenges faced which WGISS could tackle.

WGISS-57-23	TEIG to discuss a communications strategy for AI/ML white paper, to gather feedback and comments from the community. Present this to WGISS-58 for discussion.	**Due: WGISS-58**

Chat Messages

Brian Terry (SEO/AMA): Because AI/ML has limited explain ability it is crucial that the weights and training data be public if used for important decisions, for it to be reproducible

Ashutosh Gupta (ISRO): As a next step, some sample or prototype applications may be planned and hosted which demonstrate best practices from the white paper.

Brian Terry (SEO/AMA): We would like to work to make sure the example notebooks on CAL are compliant, and perhaps contribute back relevant lessons and tooling.

6.2 - Status about white paper for AI/ML

Yousake Ikehata (JAXA, TEIG Lead) reported:

− NOAA, ISRO and NASA have included their activities in the white paper.

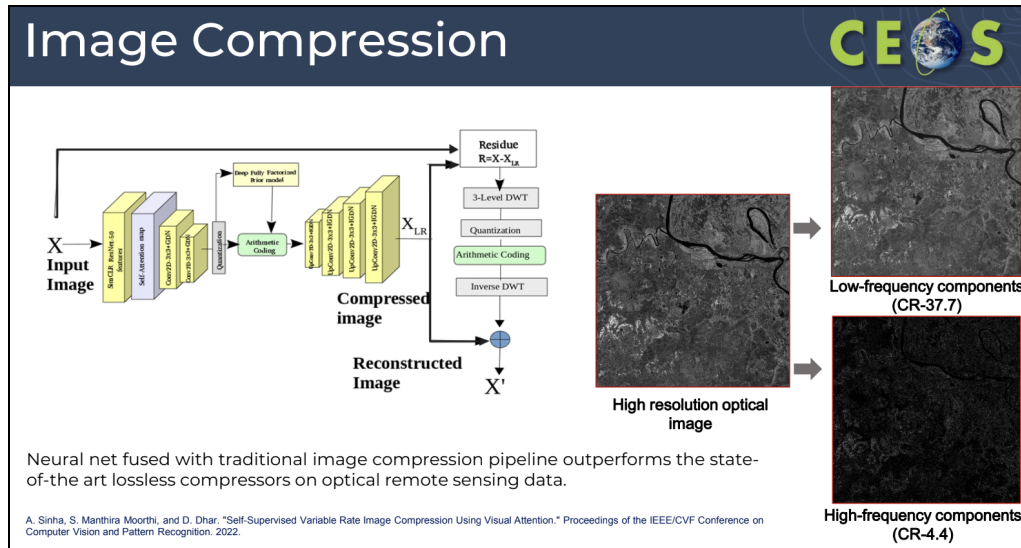− Other agencies have been invited to contribute, including NRCan, AIRCAS, ESA and INPE.

6.3 - Artificial Intelligence for Earth Observation Data Processing: ISRO Use cases.
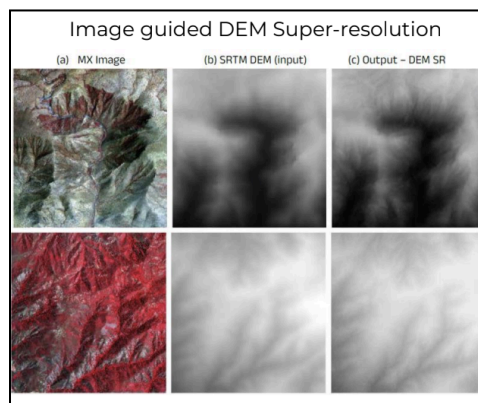
Ashutosh Gupta (ISRO) reported:

− ISRO has a large number of EO satellites across a variety of applications. They have a large group across a number of centres working on AI for EO.

ISRO – AI/ML Technology
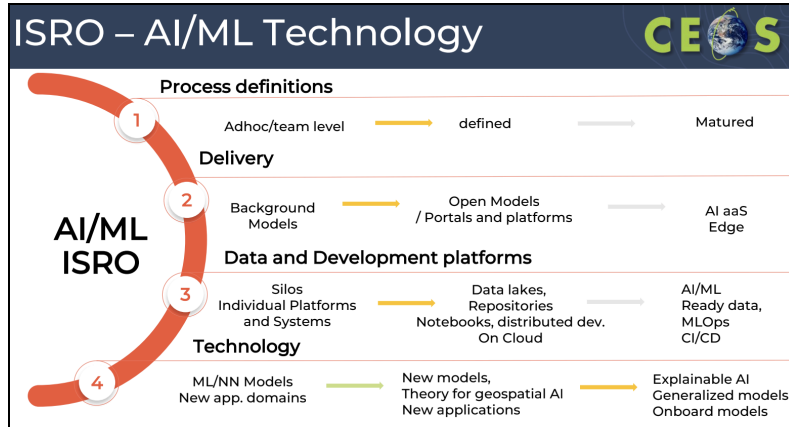
- ISRO AI work covers the whole data pipeline including acquisition, processing and augmentation and information extraction.

- ISRO has applied AI to reconstruct data where there is missing data due to a detector failure or a read-out error or transmission.

- If there is a lost multiple band sensor, ISRO has used deep learning to reconstruct the image.

- AI can be used to improve thermal band imagery. This can be used for wetland inventory and mapping, lake segmentation studies and forest fires.

- ISRO has begun using some new models such as Generative Adversarial Networks (GANs), diffusion models, and vision transformers for recovering image information. They have found the challenges to be:

  ○ Training dataset generation for supervised deep learning.

  ○ Design of network which specifically handle non-standard assumptions, e.g. non-uniform blurs, non- AWGN noise, correlated noise

Neural net fused with traditional image compression pipeline outperforms the state-of-the art lossless compressors on optical remote sensing data.

A. Sinha, S. Manthira Moorthi, and D. Dhar. "Self-Supervised Variable Rate Image Compression Using Visual Attention." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

- Images have certain features and shadows which can help improve digital elevation models.


- Map generation is a key objective of cartographic satellites. ISRO is working on image to Map translation using Generative deep learning models.

- Segmentation and classification is another application of AI/ML. ISRO has applied this to detect clouds, forest fires, snow, solar power plants and ships. Some of these models are being devised for onboard implementation.

- Object detection pipelines are also in the works. This includes content based image retrieval (CBIR) with natural language query and object indexing.

Discussion

- Yousuke Ikehata (JAXA) asked about the accessibility of the AI/ML datasets. Will it be opened in the future, noting it is restricted at the moment?

- Ashutush recognised that AI/ML data for all sensors is not identified at this stage. ISRO can let TEIG know once the dataset has been opened so it can be included in the white paper.

- Ashutush offered to help with the section on AI/ML ready data in the white paper.

- Slide 5 shows a good story for where people might use AI for EO. A similar diagram could be incorporated into the white paper.
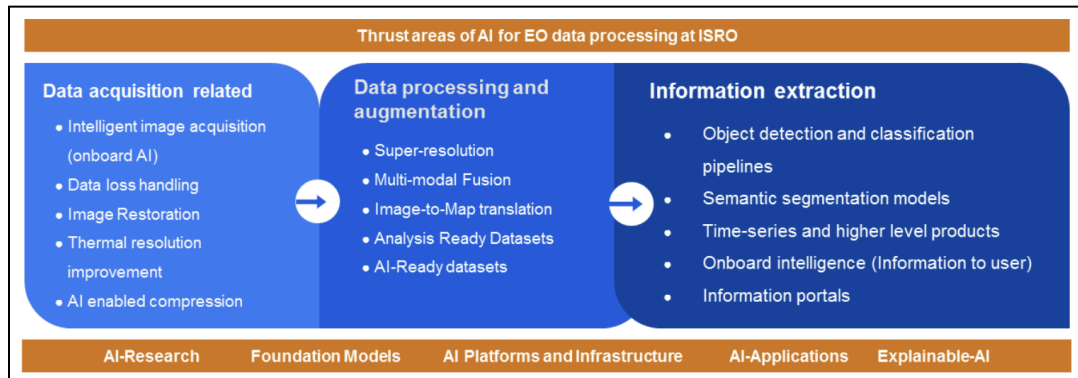
WGISS-57-24	TEIG to develop slide 5 from the ISRO presentation into a figure available on the WGISS website.	Due: WGISS-58

- The white paper should also identify and discuss major challenges and roadblocks.

- Tom Sohre (USGS, WGISS Chair) worries that we spend too long making the documents we produce perfect, when this is an evolving technology.

- Steve Covington (USGS) is interested in the image restoration application for Landsat 7. Would certainly have a mask to identify where data was made up. USGS has tried this for some forests but it wasn't great. Interested in the scientific implications and usability.

- Nitant Dube (ISRO, WGISS Vice-Chair) recognised the question for WGISS is how to tag products which have been generated using AI/ML. This could have implications in terms of disasters. How should those products be tagged?

- Ashutosh noted ISRO haven't tried Landsat 7 restoration yet - but are planning to. ISRO does hold a large Landsat 7 archive.

WGISS-57-25	ISRO to coordinate with USGS & ESA on Landsat 7 image restoration work using AI/ML to infill missing data stripes.	Due: WGISS-58

Chat Messages

Brian Terry (SEO/AMA): Because AI/ML has limited explainability it is crucial that the weights and training data be public if used for important decisions, for it to be reproducible.

Ashutosh Gupta (ISRO): Agree. In fact, explainability can be a key section in the white paper as well.

Alex Leith (SEO/Auspatious): My understanding is one of the biggest limitations in ML is the la ck of high-quality training data .

Prasanjit Dash (NOAA/COAST): How is a ML/AI based data fusion different from or compares to Kalman-filter based approaches?

Ashutosh Gupta (ISRO): The fusion here is to improve the output in terms of spatial and spectral information (pansharpening e.g.). Kalman filter fusion would typically have an identical source of information but differently prone to errors.

6.4 - CropSmart: An EO-based digital twin for making wiser cropping decisions

Liping Di (ISO/TC-211) reported:

- Goal with CropSmart is to provide farmers with the information to inform their decision making. The current approach often relies on individual experiences. The challenges involve stakeholders gaining access to timely, accurate and inexpensive decision ready information.

- Current conditions of the crop and its environment is detected via remote sensing, sensor web, IOT and modelling.

- Predicting future conditions of the crop and its environment, with and without the action relies on physical-based, statistics and machine learning models.

- Digital Twin is a digital replica of a real-world system, both the states and the functions. CropSmart Digital Twin is a digital replica of the cropping system in the Continental USA, with data of spatial resolution up to 10 m.

- CropSmart will provide details on over 10 different parameters to help farmers make decisions.

- CropSmart provides a platform for end-users to conduct trade-off studies, by providing the decision consequences through simulation for a user-given decision scenario. It will also automatically generate a decision based on the user-specified decision goal.

- The optimal irrigation scheduling tool has shown some great results for saving water.

Discussion

- Yousuke Ikehata (JAXA) asked about the time resolution for decision making. How many datasets were used to create the product? Does it rely on interoperability between those datasets?

- Liping noted EO is used to derive key parameters for each field, using Landsat datasets and Sentinel-2 primarily, alongside SMAP, MODIS and VIIRS. Machine learning based data fusion is used to fill the gaps.

- The idea is to turn this research into operational products. Want to start a company to carry the operations forward.

- Steve Covington (USGS) noted that for WGISS, it would be good to get some feedback across the search, discovery, access and exploitation of all the different EO datasets used. Always looking for good use cases.

WGISS-57-26	Liping Di to write up some feedback for CEOS Agencies about the use of EO data in CropSmart, in particular the search, discovery, access and exploitation aspects.	**Due: Q2 2024**

6.5 - Multi-Mission Algorithm and Analysis Platform (MAAP)

Doug Newman (NASA) reported:

- MAAP provides a common platform with computing capabilities co-located with data as well as a set of tools and algorithms developed to support this specific field of research. The platform addresses issues related to increased data rates and to reinforce open data policies to maximise the exploitation of EO data of the BIOMASS, GEDI and NISAR missions.

The objectives of the MAAP

Enable researchers to easily discover, process, visualize, and analyze large volumes of data from both agencies.

Provide tools and infrastructures to bring data into the same coordinate reference frame to enable comparison, analysis, data evaluation, and data generation.

Provide a version-controlled science algorithm development environment that supports tools, co-located data, and processing resources.

Address intellectual property and sharing issues related to collaborative algorithm development and sharing of data and algorithms.

MAAP High-Level Architecture

Joint Interoperability across ESA and NASA
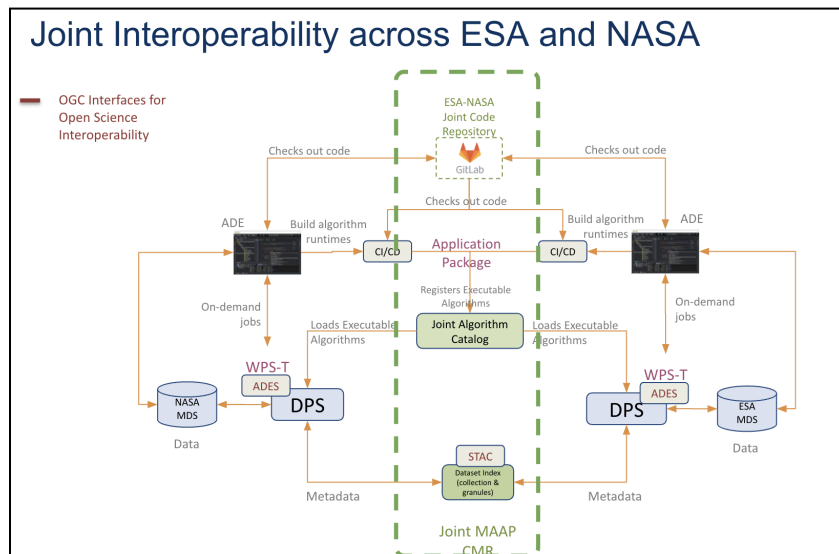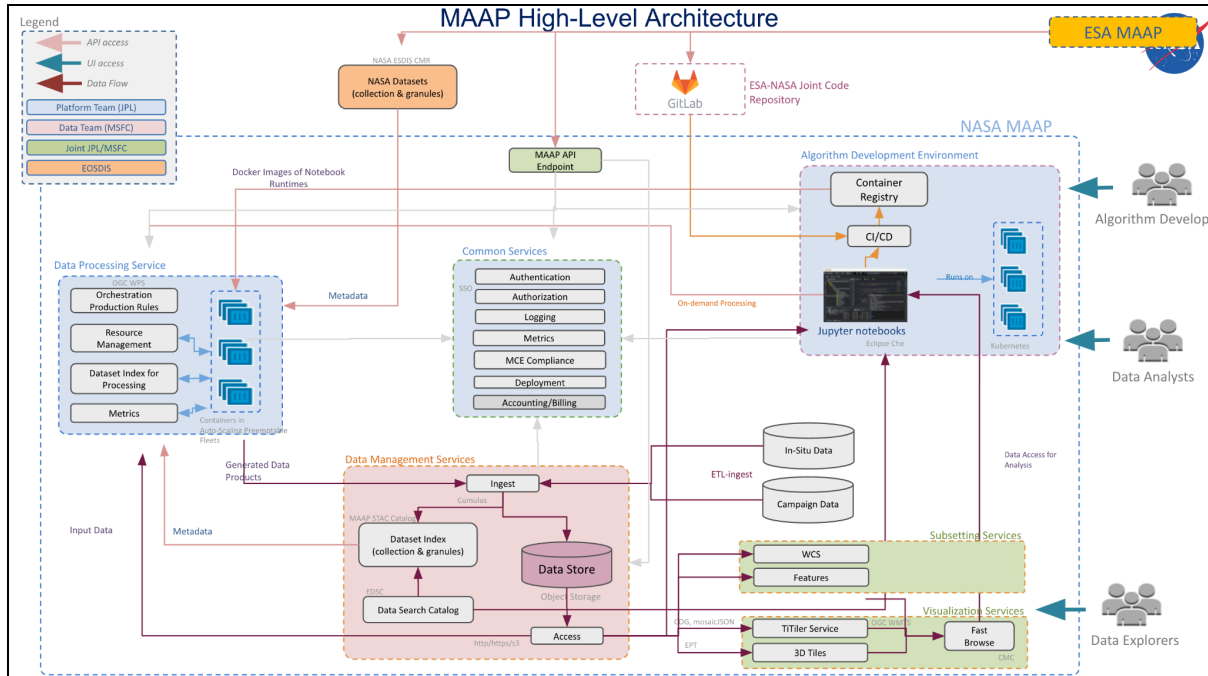
- − MAAP allows for algorithm development, data processing, data catalogues and visualisation and analysis, all in one environment, powered by open standards APIs.

- − Large scale processing allows for rapid analysis, to enhance the scientific outcomes.

- − The harmonised biomass product used MAAP to combine GEDI and ICESat-2 data. This supports UNFCCC policy reporting.

- − Dashboard was created for COP-26: https://earthdata.nasa.gov/maap-biomass

- MAAP started as a joint ESA-NASA collaboration, but has demonstrated novel approaches to interoperable and integrative algorithm development, to scalable processing, to analysis, and data exploration.

- MAAP supports NASA Earth science missions, such as NISAR, which require large-scale science analysis capabilities from L1-L3 data.

Discussion

- Egress costs are one of the biggest issues between the ESA and NASA systems. The MAAP team is looking at hybrid approaches (on-premises and cloud) to reduce costs. Not sure on what datasets will be brought back on-premise.

- Alex Leith (SEO/Auspatious) supports these sorts of environments, and is looking at the operational side of things.

- One of the use cases of MAAP was to demonstrate making the step from development to production trivial.

WGISS-57-27	Doug Newman to put Alex Leith in touch with the MAAP team for more details on the method they used to scale from a notebook to a global scale workflow.	Due: ASAP

- Yousuke Ikehata (JAXA) noted the NASA and ESA sides are mirrored. What does this imply? Doug noted the data is duplicated, but algorithms are shared. A big sticking point is the egress costs to share the data.

- Mirko Albani (ESA) noted that ESA puts data in the cloud, but not AWS. ESA can't host data outside of Europe - but it is open for anyone to make a copy and host it elsewhere. ESA data is hosted on the Copernicus Space Data Ecosystem. However, there are issues with integrity of data when a copy is made.

- Matt Paget (CSIRO) asked whether the team has considered cloud front services? Doug noted they have, but it is just another fee to pay.

6.6 - Work with Jupyter notebooks and the Digital Earth Australia sandbox

Simon Oliver (GA) reported:

- Digital Earth Australia (DEA) scientists were working independently on code with limited focus on reuse and findability. The DEA notebooks GitHub repository was created as a single development location so code could be easily shared and built upon. This evolved into a curated, mature community of practice within DEA.

- Was a useful validation tool during the development of the Open Data Cube, as well as to allow developers to communicate and test approaches.

- The DEA repository has 401 GitHub stars, and 125 code forks. It is in the top 44 of all remote sensing repositories on GitHub (out of 1612), and the top 14 of all earth observation repositories on GitHub (out of 391). 56 unique contributors across 82 Jupyter Notebooks. The notebooks are used in at least 8 university training courses, and referenced in 19 scientific papers.

- The repository is managed by a Community of Practice within DEA.

- The notebooks represent current Python libraries, and are infrastructure agnostic where possible (Desktop/Supercomputer/Commercial Cloud).

- A preferred citation is included to make it easy for people to reference the resource, and to help track uptake.

- The notebooks all use a consistent look and feel.

- GA has reviewed the current best practices draft, and noted the following:

  - DEA Notebooks follows most of the best practice recommendations, including consistent and discoverable notebook names/abstracts/headers, including products used, as well as templating and PEP8

  - GA also uses tags to aid discovery - however a challenge has been the lack of a controlled vocabulary which makes the tags less useful than they could be.

- The Sandbox is a learning and analysis environment for getting started with DEA data and the ODC. Includes sample data and jupyter notebooks that allow users to experiment with DEA's EO datasets.

- The sandbox is not a production environment.

- In the last six months, the sandbox has been logged into 5728 times, with about 30 unique users per day.

- Future work is focused on sustainability and uptake. Plan to develop a consistent guide to accessibility (e.g., colour blind friendly data visualisation), improve testing coverage of DEA notebooks to reduce manual maintenance load, and develop a scaling up mechanism.

Discussion

- The urbanisation notebook is run over one city. To run this over the whole continent, a tiled based workflow would have to be developed. To adapt, this requires quite a bit of code. Ideally users would be able to scale up without intervention from a developer.

- Matt Paget (CSIRO) recognised that when using DEA notebooks, users can explore and build on the examples. But when a user goes to scale up, they need to understand the processing workflow a bit more.

- Steve Covington (USGS) suggested that because it is a training ground, working on an algorithmic approach should be part of the training. How do people get beyond the wall of the sandbox? This should be part of the learning process.

- Yousuke Ikehata (JAXA) asked whether new training data (not necessarily EO data) could be uploaded to the sandbox? Simon noted there is limited capability to upload data, a user can only use the pre-installed data generally. Custom sandbox environments can be set up for special projects, with much larger compute, and the ability to upload external datasets.

- Matt noted that by using the STAC tools, a user can pull in any data from any source. The data doesn't necessarily have to be implemented in the sandbox, but the internet can be used to pull in any data. Anything one can do in python can be done in the jupyter environments.

- Doug Newman (NASA) asked about the strategy used to get notebooks into curriculum at universities. Simon recognised this was a gradual process, with one university learning off others, mostly through word of mouth.

- NASA is having trouble making people aware of the available tools.

- Tom Maiersperger (USGS) asked about the process to review and publish code for the default notebooks. Simon noted each contribution goes through peer review, and the community practice approach gets the notebooks to a relatively good position.

- The GitHub environment has a checklist to help ensure guide users on the best practice for contributing to the repository.

- Yousuke Ikehata (JAXA) asked about the use of the sandbox for industry and commercial work. Simon noted the computing resources are limited, but yes, it can be used for commercial applications on a small scale.

Day 3: Wednesday 6th March, 2024

Session 7: Discovery & Access

7.1 - Session Objectives

Damiano Guerrucci (ESA, DAIG Lead) reported:

−   The Discovery and Access session is broken into two sub-sessions, with STAC topics first, and the usual reporting after the break.

−   The STAC presentations have been compiled in collaboration with some of the other CEOS groups. Will conclude with a discussion about the STAC best practice document which is currently in development.

−   The two deliverables for DAIG are to finalise the STAC best practices (due in Q3 2024) and update the service discovery best practices to include STAC (due in Q3 2024).

7.2 - STAC ARD Extensions SAR and Optical

Matthias Mohr (for ESA) reported:

−   The first iteration of the STAC CEOS-ARD extensions were developed for surface reflectance, surface temperature, normalised radar backscatter and polarimetric radar. The extensions included many CARD4L fields - a term which has now been retired. The extensions were also quite complex.

−   To update the STAC extensions, a number of tasks were defined:

    ○   Create a new CEOS-ARD STAC extension, and deprecate the STAC CARD4L extension

    ○   Create specific STAC extensions for missing fields

    ○   Work with CEOS on clarifications and template(s)

    ○   Make (example) data available

    ○   Check how this relates to and can benefit the ISO/OGC ARD work

−   The CEOS-ARD STAC extension was simplified, with 'profiles' for optical and radar.

−   The extension has been extended to cater for the new PFS: aquatic reflectance, nighttime lights surface radiance, ocean radar backscatter and geocoded single look complex.

−   The optical draft is available, just waiting on a couple of clarifications. A meeting will be held next week to discuss, and all are welcome to join the conversation.

- The extension for optical makes use of the accuracy, classification, electro optical, processing, projection, raster and view geometry. It has proven difficult to accommodate per-pixel and DOI documentation required.

- The work on the radar extension has yet to start, and needs a lot of discussion for new STAC extensions. Expert assistance is needed for the SAR extension.

- Would like to try and work on profiles, which are similar to STAC extensions. Profiles are small groups of related information, which can then be chosen from to create a PFS. This would ensure no inconsistency between the PFS, and make it easier to create new PFS.

Discussion

- Tom Sohre (USGS, WGISS Chair) asked for clarification on the responsibilities for the various CEOS groups for these activities. Would like to clarify what WGISS needs to accomplish.

7.3 - OpenEO as community standard and other related OGC standards

Alexander Jacob (Eurac Research) reported:

- OpenEO goes beyond data access and discovery to processing as well. It simplified the steps to get from data access to analysis, and helps mitigate the data infrastructure and management burden.

- First step was to define a standard API to help organise the various clients, where data is then displayed in a data cube to the user.

- OpenEO supports STAC natively, and all data discovery pathways are via STAC.

- There is a long list of predefined processes included in OpenEO, to help standardise the workflows. Users can also define processes, building on existing functions, and functions, developing custom code.
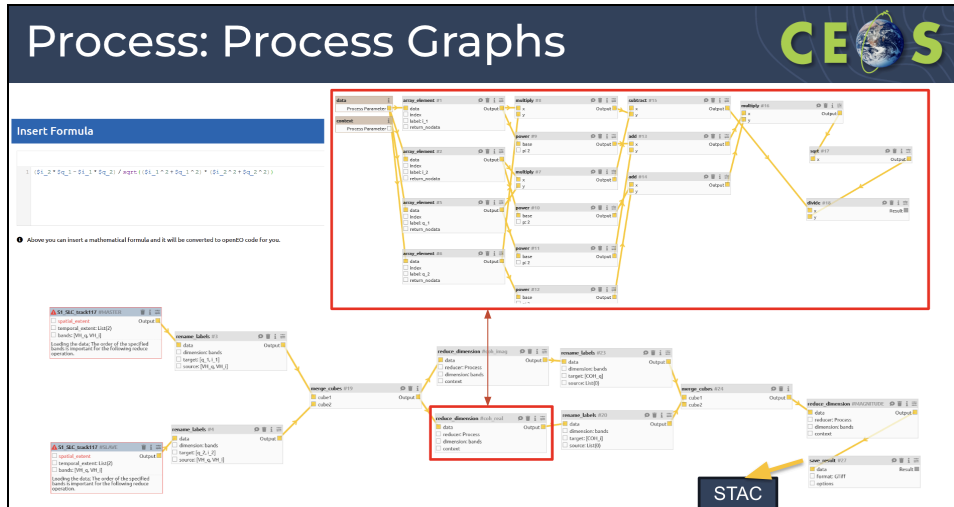
- Lots of backend clients have been built around OpenEO, Python & R.

- There is also a web editor to build processes without any code, via a graphical user interface.

- OpenEO was submitted in 2023 to OGC as a community standard. The scope covers API and Processes, and a request for comments was open throughout February. More details can be found in the justification document.

- Compared the OpenEO API to the OGC API via a crosswalk. They are similar APIs but with a different scope.

- The GeoDataCube API is an implementation profile for existing OGC APIs. OpenEO is already in many cases an implementation of GDC API. Testbed 19 covers the draft API specification, which will be produced for the OGC GeoDataCube Standard Working Group (SWG).
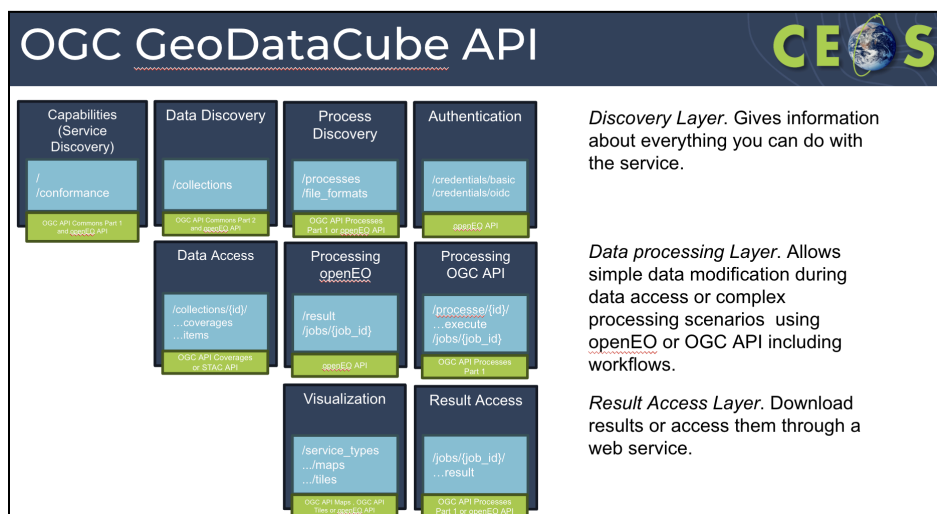
- A GUI has been developed for the GDC as well - GDC editor. An extension of the OpenEO GUI.

- OCG Testbed-19 engineering report and draft specifications were uploaded very recently to the OGC portal, and are available for review by OGC members.

Discussion

- OpenEO is an open-source software stack. Data providers are able to make use of existing software, and build on top of the data offerings.

- If USGS wanted to offer OpenEO for Landsat, they would go and grab the processes, and host that on their own cloud. Open source components to save building processes from scratch.

7.4 - STAC and STAC API Community Standard to OGC

Matt Hanson (Element84) reported:

- STAC has proven to be an interesting case for a community standard, as it has already been widely adopted by the community.

- CEOS is made up of government organisations, but in the last ten years, there has been a huge increase in industry engagement in the sector. How can we create standards for both public and private sectors?

- Industry isn't too concerned about optional standards. They care more about interoperability. In this way, there is a disconnect between government and industry.

- Specification is not a standard unless it's adopted. Creating a standard is more about driving adoption.

- Key steps to create a standard include:

  - Creating a community: gathering stakeholders, and engaging industry who will use the standard. Promoting ownership is also key, to help industry feel they are building something useful to them.

  - Leverage existing standards where possible.

  - Develop open-source implementations - not within the mission of many standards organisations. These implementations inform spec development and changes, and iterations.
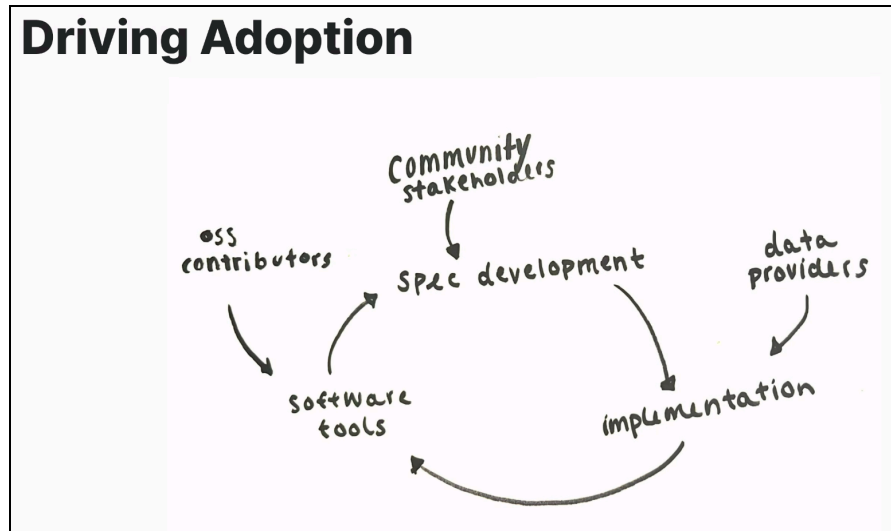
- The core specifications of STAC are simple, and are as minimal as possible to avoid frequent changes.

- STAC Extensions exist to host a marketplace of ideas, and extend STAC beyond the core specifications. There are currently over 60 extensions in total. Extensions can be rated based on their usage, and categorised into stable, candidate, pilot, proposal and deprecated.

- STAC started from implementation, with open source tooling and lots of community iteration. Early adopters helped ensure success - Landsat was one of the early adopters and really helped the community advance. Collection 2 was made available in STAC format.

- Started developing "usage" extensions, representing fields that users are not searching for, but enables connection with ODC.

- Use metadata rather than file name structure (STAC in, STAC out). Metadata contains complete characterisation of data, including data location and access.

- STAC Project Steering Committee is made up of five individuals from the community.

- STAC is hoping to be approved as an OGC community standard in April 2024. This won't change the governance of STAC.

- STAT (Spatio Temporal Asset Tasking) is a new grassroots community effort, to make tasking more interoperable based on the STAC model. A sprint is planned for April 2024 in Berlin.

7.5 - STAC Best Practices Roadmap discussion

Yves Coene (Spacebel for ESA) reported:

– CEOS OpenSearch Best Practices have been successfully used to federate and discover agency resources via IDN, CMR and FedEO.

– STAC was investigated as a successor of Open Source for resource discovery.

– There are multiple options for STAC implementation, and hence WGISS decided STAC best practices should be developed.

– The scope for the best practices is for collection and granule discovery (STAC collection/item)

– Approach decided at WGISS-55 included three steps:

    ○ Propose an applicable list of use cases to narrow scope of STAC Best Practices.

    ○ Prepare draft STAC Best Practices addressing selected use cases by solving identified issues

    ○ Obtain DAIG feedback on the draft BP and iterate.

– A survey was completed last year.

– The work will be completed in a github repository as requested by the community.

– The most requested topics were:

    ○ STAC API (75%)

    ○ Granule search (70%)

    ○ Guidance to allow federation (70%)

    ○ Advertise available queryables (65%)

    ○ Encode granule metadata (65%)

    ○ Collection metadata (60%)

    ○ Controlled keywords, media types, roles (60%)

    ○ Cloud-native assets (60%)

    ○ Collection search (55%)

– Each of these topics are addressed within the best practices.

- Proposed best practices rely on parts of STAC/STAC API, underlying OGC specifications and parts of (available) extensions of STAC and STAC API. This includes the CEOS-ARD STAC Extension.

- Asking for feedback from all WGISS members via the GitHub repository, or email.

- DAIG will plan teleconferences in April and May to discuss feedback. The goal is to consolidate an initial version of the BP in June and finalise by the end of July 2024. The finalised best practices will be presented at WGISS-58.

WGISS-57-28	DAIG to share the invite for April and May STAC Best Practice calls with WGISS members, LSI-VC leads, and ARD Oversight Group lead.	**Due: ASAP**

Discussion

- Alex Leith (SEO/Auspatious) noted that in some of the documentation, granular was mentioned. In STAC, the term 'Item' is used. It would be nice if it was aligned. Yves noted that the term granular was used based on previous WGISS documentation, as this is a WGISS document.

- Doug Newman (NASA) recognised more contributions are needed, and Alex is welcome to propose changes on the GitHub.

- Doug noted that CWIC allowed for search of a catalogue, and to do file/granular level searches within the collection. That search was then federated out to the OpenSearch API to connect to other agencies. This was pretty powerful, but OpenSearch wasn't well adopted. APIs started to rot and become deprecated - this is an opportunity for STAC.

- Tom Sohre (USGS, WGISS Chair) suggested that a self-compliance assessment could be set up. Doug agreed this should be encouraged, with the benefit being that the data is more usable. Giving visibility would also be beneficial, showing who is and isn't compliant.

- The team would like more people reviewing and providing comments in the best practices repository.

- Filippo Marchesi (ESA/Solenix) has been attending various CEOS STAC related meetings, and has pushed this document quite a bit but hasn't seen much response.

- It is important WGISS confirm with the various CEOS & external groups to ensure they are happy with the best practices before they are endorsed as version 1.

- Damiano Guerrucci (ESA) recognised that if a CEOS member wants to implement STAC, they could be confused by the number of conversations. The goal is to start from the basics with this paper, without entering too much into the technical elements. By setting this basis, this should give data providers an indication of where to start.

- Tom noted that perhaps the name of the document itself can be confusing, as there are different perspectives on what STAC means. Should it be called "Collection and Granule discovery BP with STAC"?

- The document should be sent to all data providers within CEOS Agencies, to assess their practices against.

- Damiano recognised that what WGISS does is not going to change what is happening in the STAC community. Even if STAC is accepted as an OGC community standard, the governance won't be changed. From the CEOS side, WGISS should focus on giving guidance on how to enter into the domain.

- Yves noted that if the STAC steering committee could approve the CEOS STAC best practices, that would be nice, but not sure this would be possible.

- Nitant Dube (ISRO, WGISS Vice-Chair) recognised it is easier to generate best practices for CEOS rather than the whole community.

- Alex noted the datacube point for STAC is key. Great to have a good metadata catalogue, but being able to download the items is key. Some key fields within the extensions make it possible to be highly efficient.

- Doug suggested WGISS should expand on the scope once this best practice is completed. The sooner this document is in the hands of people, it will make producing the following documents easier.

- Yves suggested the title could be "Collection and Granule discovery with STAC Best Practices". If the STAC community wanted to adopt some of the items into their own STAC Best Practices, that would be great and CEOS can then point to something.

WGISS-57-29	DAIG to consider updating the title of the STAC Best Practices to "Collection and Granule Discovery Best Practices with STAC", and proceed with Best Practices finalisation according to the plan presented by Yves Coene.	**Due: WGISS-58**

- Alexander Jacob (Eurac Research) noted the document is very data provider centric, but WGISS should also consider users to some extent and the processing back end.

− Richard Moreno (CNES) recognised this is a good topic for the federation activity. OpenEO is a good example of managing a federation with different copies of data.

− Nitant recommended the interoperability handbook reference these best practices.

− Matthias Mohr (for ESA) recognised there are some items which are specific to the CEOS community and not the larger STAC community. But there would be good inputs to the STAC community.

Chat Messages

Alek Cesarz (CDSE): Assuming GDC will be accepted as standard, how likely is it that, with the current openEO backend, providers will move to GDC?

Alexander Jacob (Eurac Research): First of all openEO is already largely following OGC API in terms of the API spec, GDC from openEO existing providers, can be seen as an extension and should be relatively easy to add into existing services, if they want to have the ability to add application packages as processes into their backend.

Yousuke Ikehata (JAXA): After publishing OGC API, can we get "federation" by implementing OGC API for each system? I can't understand the difference between OGC API and OGC GDC API. OGC GDC API includes the whole element; authentication, data access, processing, visualisation and so on.

Alexander Jacob (Eurac Research): OGC API is a collection of standards defining different capabilities. So it actually is not one API but 15 different ones. GDC API creates a profile selecting many of these different building blocks and combines them into one API.

Alek Cesarz (CDSE): Is it possible for us (CDSE) to participate in the discussions on the STAC best practices? We are in the middle of reengineering of the STAC API of CDSE and face multiple decisions which we would like to base on what is planned to be the industry best practice.

Alek Cesarz (CDSE): STAC offers a lot of flexibility in collection/item definition. This is great. but for the sake of interoperability of instruments and products the requirements for data providers should be quite strict. and synced with the STAC development.

Matthias Mohr (for ESA): I think the STAC community needs to guide people better, what extensions to use in which cases, etc. Maybe with profiles similar to the STAC CEOS ARD profiles I presented earlier, but simpler.

Alek Cesarz (CDSE): This guidance could come from CEOS Best Practices and the work on profiles. If CEOS Best Practices covers this part and is advertised through the STAC community then it would also reach New Space.

Matthias Mohr (for ESA): But that should be developed in the STAC community together with CEOS, not by CEOS without the STAC community.

Peter Strobl (EC-JRC): My understanding of CEOS-ARD would be exactly that. It should be a blueprint to present data such that they are interoperable by design (at least within a defined 'domain' or 'level').

Matthias Mohr (for ESA): Yes, although the CEOS ARD seems to be too complex. The best practices must be somewhat simpler.

Matthias Mohr (for ESA): The data cube extension in openEO is required because openEO works on top of data cubes. It's needed for processing.

Peter Strobl (EC-JRC): Yes, but unfortunately there is no common (beyond openEO) definition of what a datacube is (and what is not).

Matthias Mohr (for ESA): Data cube is specifically needed for openEO as it always creates data cubes on the server-side, but for most general consumption uses raster, proj and potentially eo/sar/etc. are needed. A couple of others for certain use cases. And to explain what is useful for which use cases is important.

Alek Cesarz (CDSE): The data distribution and processing platforms, like CDSE are in the middle and need to follow both the provider side of specs to be able to understand and catalogue the data, and the user side as well to be able to serve the data in a unified way. This drives the need for alignment of the specs for things critical as metadata.

**7.7 - FedEO Report**

Yves Coene (ESA/Spacebel) reported:

− FedEO is part of the WGISS Connected Data Assets. FedEO Client is available at https://fedeo-client.ceos.org, and the FedEO Server (API) is available at https://fedeo.ceos.org/readme.html

− The validation server (API) is available at https://geo.spacebel.be/readme.html

− New FedEO version was released recently, and includes RESTful API offering multiple interfaces including OpenSearch, OGC API and STAC API.

- FedEO has grown from 172 to 223 million granules (+29%) since WGISS-56 in October 2023.

- ESA (EOCAT) contains 182 collections with 34.9 million granules.
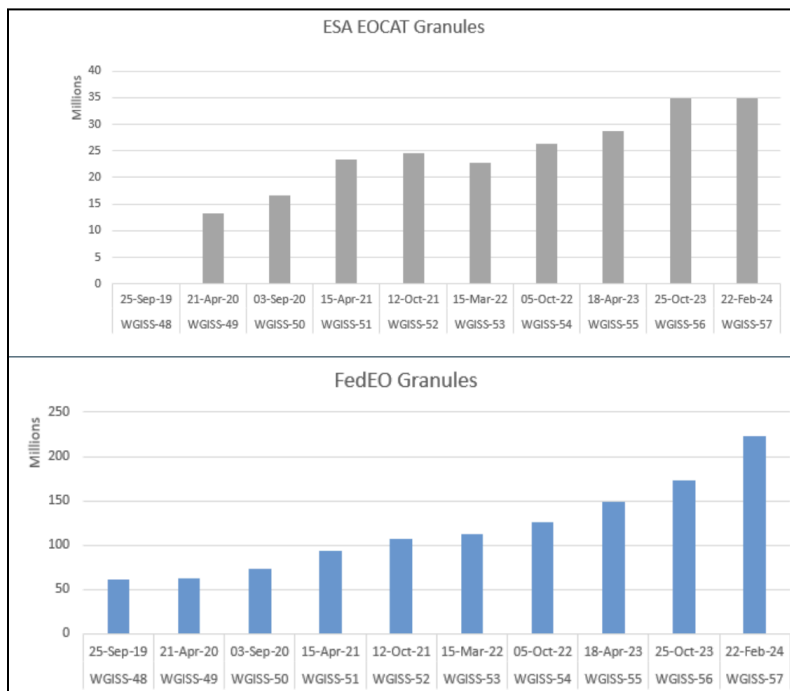
- A periodic notebook metadata record harvester has been developed, which uses GitHub/GitLab API to find all notebooks, extracts embedded EUMETSAT-like metadata and accesses DataCite to obtain citation. This has been applied to 4 GitLab/GitHub repositories.

- New FedEO API also supports service discovery, including tools from ESA Earth Online.

- The STAC encoding on FedEO is also being improved.

Discussion

- Tom Sohre (USGS, WGISS Chair) suggested an announcement be made through the CEOS communications channels for the release of the new version of FedEO.

WGISS-57-30	DAIG to work with the CEOS Comms Team to put together an article for ceos.org/news regarding the new FedEO release and the retirement of CWIC.	**Due: Q2 2024**

7.8 - International Directory Network (IDN) Report

Michael Morahan (NASA) reported:

- IDN Search Portal contains all dataset metadata entries in the CMR.

- CWIC will be retiring this spring - by the next WGISS meeting, CWIC will be dropped.
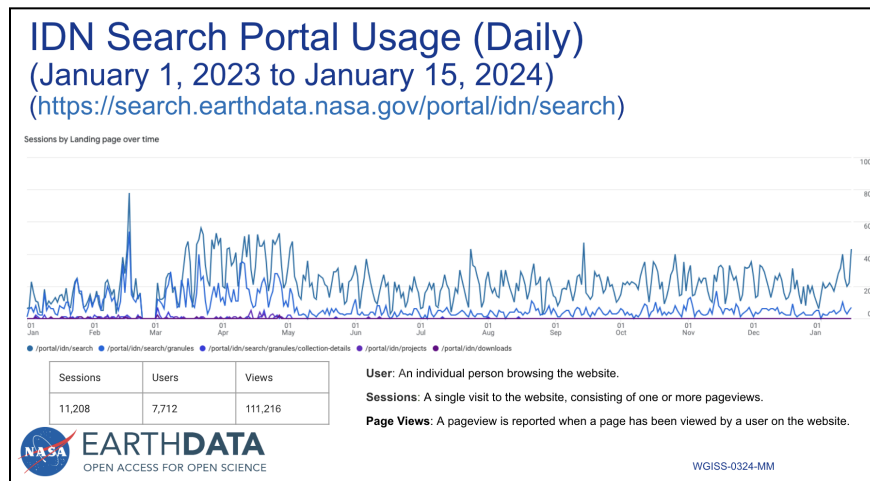
IDN/CWIC Metrics

- Total Collections and Granules offered by the CWIC and FedEO:

CEOS Interfaces	Total Collections	Total Granules
CWIC	~369	~17,811,018
FedEO	~510	~60,846,230
JAXA/FedEO	1017	~13,740,118

- Total Collections and Granules offered by the IDN:

IDN Search Portal	Total Collections	Total Granules
IDN	~52,985	~1,401,651,712

- IDN has tested the new FedEO server for ingest of ESA, ESA CCI, VITO, DLR and JAXA collections, granule discovery and download.

- Performed an ingest test of NOAA/NCEI collections, with a total number of records 37,335 and 24,873 ingested into CMR.

- With ISRO, the existing collection metadata records have been updated, and the download links for CEOS 2-step collection-granule discovery and download through OpenSearch APIs have been reviewed.

- IDN will review INPE STAC API and clients.

- A number of updates were made to the IDN and CWIC pages on the CEOS website. The CWIC page will become more of a description about CMR.

- The ceos.org IDN page will be retired, and moved to an NASA Earth Data page. There won't be a permanent redirect in place. An email was sent to the user base.

WGISS-57-31	WGISS chair to work with SEO regarding the retirement of the website subdomain for CEOS IDN, and forwarding to the NASA IDN page.	**Due: Q2 2024**

- The new IDN homepage will link to CEOS specific queries on the IDN Search Portal, including searches on keywords provided by CEOS MIM Database.

- Links to important sites will be included: Unified Metadata Model (UMM), Directory Interchange Format (DIF) Standard and Global Change Master Directory (GCMD) Keywords site.

- A number of new GCMD keywords have been added as well.

- The metadata management tool (MMT) has had six unique authors use the tool from August 15 2023 to Jan 31 2024.

Discussion

- CMR provides all the CEOS Member agencies datasets. Are there contacts for each of the agencies? Tom Sohre (USGS, WGISS Chair) was unable to find Landsat data in IDN.

- Michael noted USGS EROS and LTA data is being ingested by someone at PDAC, but Tom doesn't believe this is occurring correctly.

WGISS-57-32	IDN team to review their POC list with respect to the list of WGISS principals, and confirm their representation in IDN. In particular, confirm the USGS contact. Set up a process to regularly (annually) ensure the POC is up to date, and any agency datasets represented in IDN are up to date.	Due: Q2 2024

- Yves Coene (ESA/Spacebel) noted that FedEO gathers the information automatically from the agencies.

- Nitant Dube (ISRO, WGISS Vice-Chair) recognised there is an automatic process for JAXA data to be ingested into FedEO. With the new data policy announced from ISRO, they need to add a lot of new collections from ISRO. Is there a way this process can also be automated for ISRO?

- Doug Newman (NASA) noted this can't be done yet, but the team are hoping to have some less stringent authentication issues soon. The team will do their best to improve the process, but there are some policy problems to fix first.

7.9 - CEOS Data Access Website

Filippo Marchesi (ESA/Solenix) reported:

- The group was previously referred to as "System Level Team" (SLT) which was renamed as Data Discovery & Access Interest Group (DAIG).

- DAIG webpage has been updated with new content and links.

- The mailing list access-systeam-help@wgiss.ceos.org directs to Damiano, Filippo, Yves and Michael.

- The procedure to add data to the CEOS connected data assets programme is manual at the moment, via an email and excel spreadsheet. Would like to make it more dynamic in the future.

- The CWIC page will go through some significant updates in the near future.

- The open source software page was removed from the website, and the content was migrated to appropriate pages.

Discussion

- NASA isn't listed on the graphic shown. NASA data is already within the CMR system, so doesn't technically fall within this graphic, but it probably should be added for completeness.

|		
WGISS-57-33	DAIG to add NASA on the connected data assets graphic, and strengthen the connection between IDN and FedEO. Potentially also add other WGISS "members", even if they are not currently uploading metadata to FedEO or IDN.	**Due: ASAP**

- Iolanda Maggio (ESA/Rhea) noted the WGISS-52 action on open source software inventory. A new page has been created, does the old page need to be removed?

WGISS-57-34	DSIG to confirm with SEO to take down the Open Source Software Inventory, and make sure it is removed from the CEOS Website.	**Due: ASAP**

7.10 - CEOS Jupyter Notebook inclusion into the service catalog – EUMETSAT use case

Yves Coene (ESA/SpaceBel) reported:

- CEOS Best Practices for Service Metadata and Discovery version 1.0 was approved and published in November 2022. The discovery scenarios apply also to Jupyter Notebooks with downloadable tools, Web GUI tools and coupled services.

- EUMETSAT have notebooks publicly available for training purposes. Have demonstrated the discovery and access of these notebooks via FedEO.

- FedEO Browse tools organise the STAC catalogs by collections as well as services and tools. It assumes the resource metadata is annotated with controlled SKOS keywords/concepts.

- Notebook metadata is extracted via GitHub API, and is enriched with text content, including DOIs and keywords.

Notebook Discovery – Metadata Harvesting (3/3)
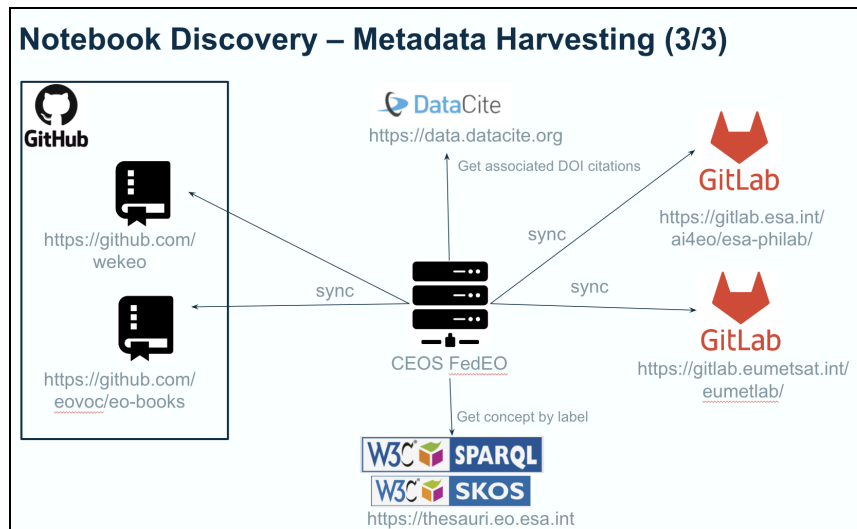
- Notebook metadata can be accessed via static STAC catalogs, STAC API or OpenSearch.

- Can resynchronise every 24 hrs if desired, but are unsure at this stage how this would scale.

- To be eligible, the notebooks would have to meet the metadata criteria.

- Can also use STAC API search requests to find tools and notebooks about a specific topic.

- Propose that forthcoming "CEOS Jupyter Notebook Best Practice" recommends embedded Notebook metadata, and recommends controlled keywords (vocabularies) to be used: platforms, instruments, science, … (e.g. GCMD).

Discussion

- The EUMETSAT notebooks have a good style, and hope that they can be part of the best practices to use as a style guide.

WGISS-57-35	TEIG to incorporate the recommendations from Yves in the Jupyter Notebook best practices.	Due: ASAP

- To add other notebooks, agencies can reach out to Damiano.

- Perhaps the notebooks should be included in a diagram to accompany the connected data assets diagram.

- Is there a metadata tag which indicates the peer review status?

- Matt Paget (CSIRO) wasn't clear how to proceed on this topic after the last WGISS. Generic notebooks are a bit difficult, as they are built to run on specific environments.

For example, GA notebooks are set up specifically to run within their environment. The code is general, but couldn't run in a generic environment.

- Discoverability is the goal, at least for now. Will have to investigate how the notebooks should be executed.

- There could be a metadata field for which environment the notebooks are written for.

Session 8: Data Preservation and Stewardship

8.1 - Work Plan for Data Preservation and Stewardship Interest Group (DSIG)

Mirko Albani (ESA, DSIG Lead) reported:

- The activities of DSIG cover:

  - The sharing of agency investigations, developments, experiences and lessons learned relating to EO data stewardship.

  - Development of common cross-agency best practices or guidelines of data stewardship for possible adoption by CEOS/WGISS members.

  - Sponsor technical exchanges and the conduction of joint activities and/or pilot projects on specific data stewardship topics.

  - Establish and maintain a CEOS "Data Purge Alert" service.

  - Contributing to GEO and standardisation activities.

- Several best practices have been written by DSIG. Some key outputs by the group include:

  - Data Stewardship Reference Model

  - Data Management and Stewardship Maturity Matrix (DMSMM)

  - Archive Technology Trends and Evolution

  - Purge Alert Service White Paper

  - Heritage Data Recovery

- Both 2023 work plan items for DSIG were closed on time.

- For the 2024-2026 CEOS work plan, the proposed deliverables include:

  - White Paper on EO Data Collections Management and Governance

  - White Paper on Software Preservation

Discussion

- Tom Sohre (USGS, WGISS Chair) asked about the definitions of the terms white paper and best practice. Have we clarified the difference?

- Iolanda Maggio (ESA/Rhea) noted that a white paper should explain a technical topic. A best practice is a delta on top of this.

- Tom added that WGISS should make sure that when a document is published, it is promoted in some way.

- Bringing to plenary for endorsement can bring more attention to a specific document, but including it in the annual WGISS Chair report would probably be sufficient.

|		
WGISS-57-36	WGISS Chair to work with CEOS Executive Officer regarding document governance including definitions for best practice vs. white paper vs. case study, and at what level (WG vs. SEC/Plenary) each type of document should be approved.	**Due: WGISS-58**

- Nitant Dube (ISRO, WGISS Vice-Chair) asked how DSIG plans to contribute to the interoperability handbook. Policy and architecture would be the most relevant places to contribute.

8.2 - Archive technologies evolution white paper: summary and recommendations

Iolanda Maggio (ESA/Rhea) reported:

- Archived data volume at space agencies is growing exponentially. At the same time archiving technology is changing very rapidly.

- Surveys, workshops and dedicated sessions were held by WGISS to present the status of CEOS agencies archives in terms of content, technology and future evolutions/trends. A dedicated study on the evolution of the ESRIN archive was also performed at ESA.

- These activities were summarised in the CEOS Archive Technology Evolution White Paper.

- Long term archive technical meeting was held at ESRIN in 2023, to understand the challenges faced by different agencies.

- The ESA/ESRIN Archive Evolution Study aimed at drawing a roadmap for the update of the ESA Heritage Space Data Archive, has been used as an input for the white paper and covered the following topics:

  ○ Archiving market (hardware and software)

  ○ Cloud providers (on premise, public offers, etc.)

- ○ Standards

- ○ Existing ESA technologies and archives solutions (Copernicus LTA, planetary science Archive, etc.)

- ○ Other organisations technologies and archives solutions (NASA, CERN, NOAA – NCEI, EUMETSAT etc.)

- ○ Sustainability (energy consumption, cooling requirements etc)

- ○ Costs (HW, SW, maintenance etc.)

- – The white paper takes the ESRIN archive as reference and describes the functionalities and architecture of the archive, which is similar to most of other solutions implemented at WGISS Agencies. It collects some of the main challenges faced today by EO data archives, provides information about archiving and preservation solutions available on the market, and provides recommendations for archives evolution taking into account environmental friendly aspects (power consumption).

- – Challenges involved in using tap-based archives include:

  - ○ Data migration is complicated, time consuming and expensive

  - ○ Longevity of technology has side-effects (maintenance costs, obsolescence of HW)

  - ○ Tape prices can be unpredictable

  - ○ With the end of some tape families and HSM archive managers, it is feared that the main vendors are moving away from tape archive solutions

  - ○ Archives are ready for lots of data, but less ready for billions of small files

  - ○ Extracting data from tapes is slow and complicated

- – There are however lots of advantages of using Tape Based Archive, including, easy to expand, power efficient, tapes are cheap, tapes evolution are on par with disk evolution in terms of density, and malicious encryption or deletion of data is just not possible.
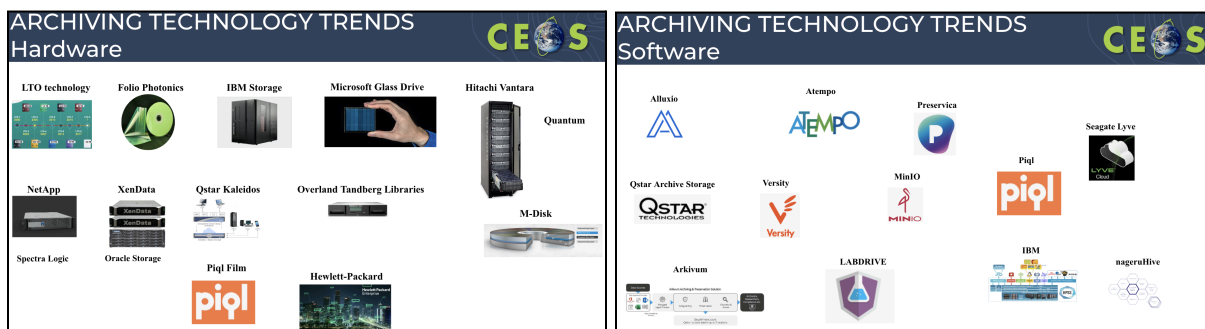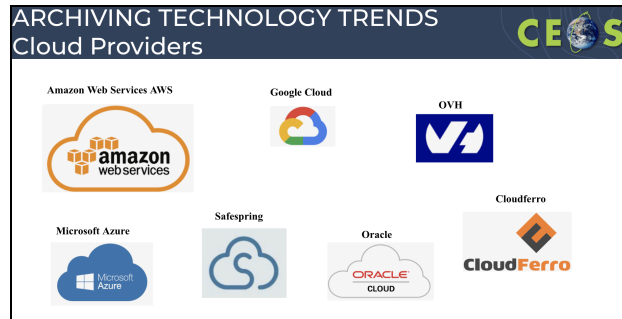
- There are solutions to keep using currently owned storage technologies, without needing to change the entire infrastructure.

- Hardware providers are developing hot, warm and cold storage solutions that can complement the current infrastructure.

- Some WORM storage options are already available or being developed that are based on nonmagnetic media, like optical disks or crystal based.

- More organisations are moving to cloud-based services because of their flexibility and cost effectiveness in the short-term. However, the step to cloud migration needs to be planned because long-term costs and vendor locking could be an issue in the future.

- Complementary to pure cloud solutions, the hybrid-cloud approach is suggested. A mix between public access on the cloud plus cold on-premise storage would provide benefits in the long-term.

- The preservation software market is very active and growing. In addition to the classic providers in this area, some new players are showing interesting solutions also with a very open-minded approach to avoid extra costs or cloud locking.

- Software-as-a-Service (SaaS) is a trend in this market.

- Many providers are showing some environmentally friendly messages, it is however not clear how much of the figures are greenwashed.

- Some high level recommendations are:

  ○ To ease data migration, vendor lock for both hardware and software is to be avoided when possible.

  ○ To ensure an environmentally friendly evolution, tape based solutions must be taken into account.

  ○ Complementary to pure cloud solutions, the hybrid-cloud approach is suggestable.

  ○ OAIS is the standard for preservation and needs to be followed.

○ Big cloud providers are charging for requests or operations which make it almost impossible to make a budget in advance.

○ Some European cloud providers are not charging for the recovery.

Discussion

− Steve Covington (USGS) noted that often policy can't be separated from preservation technology. If all reproducible datasets are eliminated, the archive becomes more manageable when looking at the technology and cost efficiency. For example, the archive of irreproducible Landsat mission data over 50 years is only 3 PB rather than 20 PB.

− Depends whether it is cheaper to reprocess or store the higher level products.

− Mirko Albani (ESA) noted that for higher level products, it is unclear how many copies need to be kept in the archive. The raw satellite data is the most important.

− Policies weren't included in the archive technologies white paper. The CEOS recommendation is to keep 3 different copies of the data in different locations.

− Tom Sohre (USGS, WGISS Chair) recognised many different organisations are looking at the policies about what they really have to do regarding their archives. It would be nice to see how the different agencies handle this, not to agree, but just to understand.

− Perhaps producing a best practices for archive technologies could be a next step.

WGISS-57-37	WGISS Exec to review the [New Agency Dataset page](#), and whether it is still used. The process for updates to this page should be clarified.	**Due: WGISS-58**
WGISS-57-38	DSIG to schedule a discussion for WGISS-58 about the various agency archiving policies. Consider developing a best practices document for archiving policies (and technologies).	**Due: WGISS-58**

− Steve asked about the pros and cons of a SSD versus a tape? SSD are getting cheaper, and are just as green as tapes.

− Doug Newman (NASA) noted some quick research indicated tapes are much more carbon-efficient to produce than SSDs, but requires more detailed research.

− Steve recognised the cloud is enabling long term analysis, so don't necessarily want the data to sleep.

Decision 01	The Archive Technologies White Paper was endorsed by WGISS-57.

8.3 - Long Term Archive Workshop

Iolanda Maggio (ESA/Rhea) reported:

- The main challenge of archives is the sheer volume of data, the copious number of files, the diversity of formats and the yearly increase of data produced.

- The archives are constantly enhanced to facilitate the archival process by using technological innovation and practices. To save energy, thus making the archive and its hardware environmentally friendly, green computing technologies are often favourable.

- A summary of each data provider's challenges and status was presented, including from QStar, Pure Storage, Spectra Logic, Quantum, PIQL, Dell and NetApp.

- The reasoning for using tape storage on-premises was discussed, highlighting that generally no external access is required. EO data records are the main asset of the ESA EOP department, and having a copy on-premises is very important. The choice of using tapes is also based on the fact that tapes have been used for years and have proven to be reliable and have advantages (e.g. cost, security, environmental) when storing multiple copies of the data.

- There is a need to evolve into a "warmer" storage solution in the future, perhaps by following the example of what DLR has done, where they went from a cold archive to an archive that works with HPC. So, the data is exploited in the layer on top of the archives.

- HPC (High-Performance Computing) solutions are under consideration, in particular for frequently accessed archives. To feed HPC, one needs high-performance data as well.

- The difference between the dissemination systems and storage was raised. These are separate environments, but an approach is needed to optimise which data is stored and which is on the dissemination systems. Not all data is wanted equally, and different "tiers" of storage may be the best approach.

- The issue of data duplication and integrity was also raised, highlighting that copies may be spread everywhere on the cloud, and how to ensure that a user gets the original data not manipulated.

- The merits of using a hybrid system of proprietary and open storage from a security point of view were discussed. In many cases, clients will employ a double copy of data, with one utilising proprietary technology (e.g. for enhanced security, or improved

performance) and the other with an open standard (e.g. LTFS, to ensure persistence of access in the future).

− It was agreed to organise another workshop in a couple of years time.

8.4 - WGISS documents and past activities pages periodic review: proposal for discussion

Iolanda Maggio (ESA/Rhea) reported:

− WGISS web pages should be reviewed every two years.

− New 'Best Practices and Guides' page was published. The content is separated out by interest group, and includes a review date column. Interest group leads should go through their documents to make sure they are up to date on a regular basis.

− At the bottom of the document page there are a few documents listed under 'Document Archive', which contains documents which are no longer relevant.
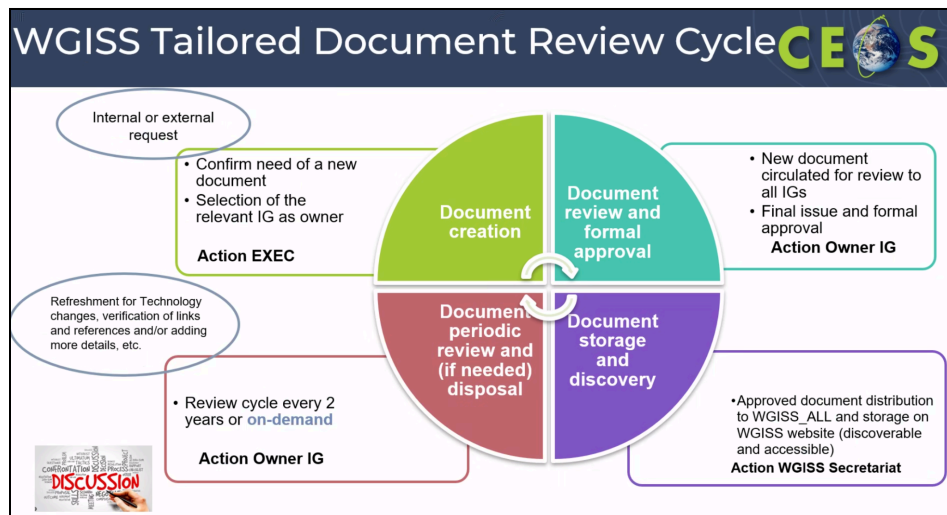
− Existing documents from other sources may be of interest to WGISS, and should be included on the WGISS website. The proposed process for these documents are:
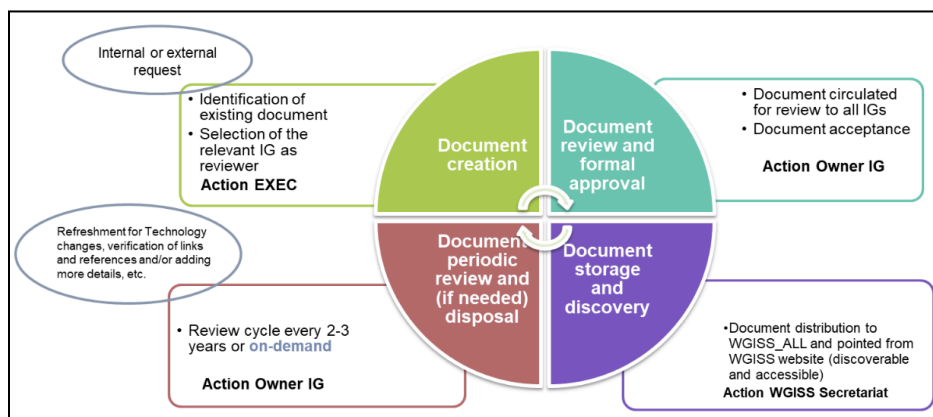
- Plan is to include a new section on the documents page.

- Suggest to rename past activities page to 'collaborations' - and differentiate between ongoing and completed activities.

- The Heritage Data Recovery section will be added in the DSIG section, including the AVHRR work.

WGISS-57-39	Libby to work with DSIG to draft up the collaborations page on the WGISS website.	Due: ASAP

8.7 - CSIRO AVHRR activity update

Matt Paget (CSIRO) reported:

- Edward King gave a presentation at the last WGISS meeting on CSIRO's AVHRR recovery efforts.

- AVHRR records span from 1992-present, with some potential to restore data from the 1980s as well.

- The higher temporal resolution of AVHRR over Landsat allows for more detailed analysis.

- In the last 6 months, the focus has been on three areas:

  ○ Trial L1 processing of existing (stitched) archive, with 61022 stitched passes from 1992/04/01 – 2023/12/31 collected from daytime measurements from NOAA-11/14/16/18/19.

  ○ Recovery of pre-1992 archive: continue sorting through data (on disk), and read back from 10,000+ tapes. Have identified pre-1992 data from Tasmania, Queensland and Western Australia. Working on reverse engineering data file formats (station specific), with the plan to extend basic processing to the 1980's.

  ○ Data availability: working to make data (L0 + L1) publicly accessible through CSIRO research data portal, EASI cloud, and TERN Data Portal.

- CSIRO is still exploring and developing the archive, checking the data for inconsistencies.

- The data is publicly available on EASI, but there are some issues to work through.

- Have also started exploring the atmospheric correction of AVHRR top of atmosphere.

- For radiative transfer modelling, CSIRO are evaluating the suitability of RTTOV for atmospheric correction. Other RTMs being considered include 6SV, libRadtran and MODTRAN6.
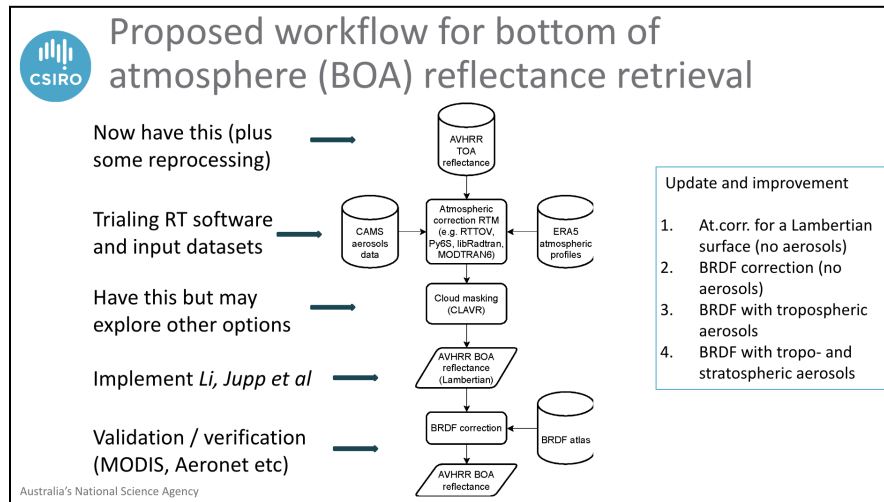
- Current work CSIRO is doing is recognising the archival work done by many partner organisations over the last 40 years.

- Hope to have it online and ready to use this year.

Discussion

- Tom Sohre (USGS, WGISS Chair) asked whether this data would show up in CMR or FedEO? Matt confirmed that it would. DSIG should eventually make all the recovered AVHRR data available in one place.

- Mirko Albani (ESA) suggested developing a 'super-collection' which brings together the different archives managed by the various agencies. This would be down the track, and would need to consider data formats and ensure consistency in processing.

- Mirko is working with CSIRO, and is interested in how CSIRO are stitching the data, because ESA are not unable to unstitch their archives. ESA would also like to share the Antarctica archives with CSIRO, and aim to harmonise what the two agencies are doing.

- Nitant Dube (ISRO, WGISS Vice-Chair) asked whether the different software could be shared via a GitHub repository or other environment. Could that help in getting engagement and consistent products?

8.6 - AVHRR Data Recovery Project Updates

Mirko Albani (ESA) reported:

- AVHRR data is available as Global Area Coverage (GAC) (4 km) or Local Area Coverage (LAC) (1km). 1km AVHRR LAC data gives new insights on structural changes on land in comparison to coarse resolution AVHRR GAC data.

- The NOAA GAC Archive covers from 1978 onwards. But for LAC data, it is only globally available from MetOp AVHRR instruments flown from 2008 onwards. Many national / regional data archives of LAC data exist, which cover a longer period. This has high value for the retrieval of Essential Climate Variables (ECVs). Some of the archives are accessible for users, others not due to unknown accessibility, responsibility, data format and structure.

- There was a request from the climate research community to have better access to these long time series.

- The objective is to build a worldwide coverage AVHRR LAC data series from 1979 onwards. The tasks include:

  ○ Unfolding and making accessible 1 km AVHRR data from regional archives (possibly open and free);

  ○ Transcribing unique data from heritage media;

  ○ Identifying a common format for AVHRR L1B and L1C data and pursuing (re)processing from AVHRR data owners/holders and data accessibility;

  ○ Facilitating data discovery through the WGISS Connected Data Assets Infrastructure.

- Two different global Land 1-km AVHRR datasets covering the period 1992-1999 are available at ESA:

  ○ Data acquired at ESA network stations (Terranova, Nairobi, Manila, etc.) were processed up to L1C and published in the ESA web page as dataset Out-of-Europe.

  ○ Data in stitched format (.arch files) from USGS, but are not processable at ESA (and so not accessible to users) as data is not congruent with the expected one.

- ESA is looking for documentation on the stitched product formats. Have tried to find information from USGS or NOAA, but didn't succeed. Not sure how to proceed further. CSIRO may be able to help with the methodology.

- For the regional archive, ESA processed the time series from 1981 - 2000 for Europe. Unique source to retrieve Essential Climate Variables (ECV) to investigate climate change over the last 40 years.
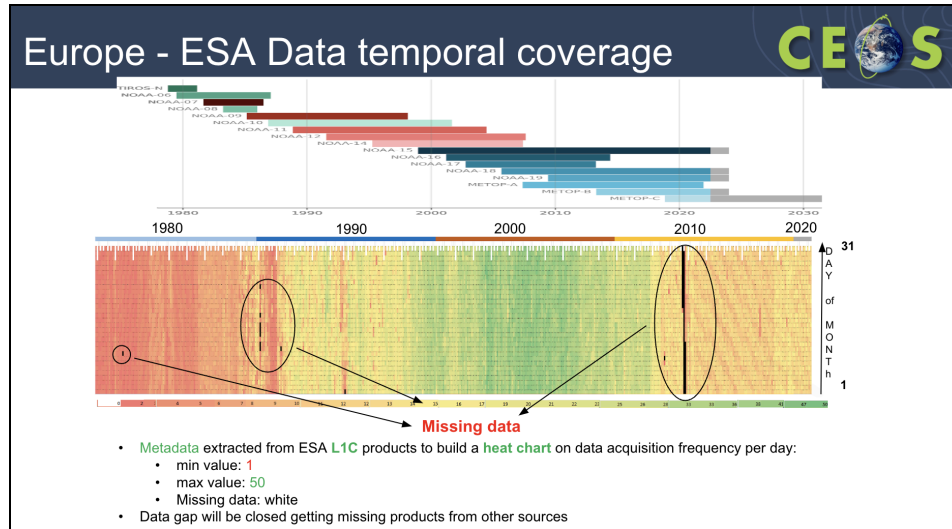
- ESA has contacted the Polish Institute of Geodesy and Cartography regarding their archive from 1996. These products aren't available online. Will hope to have a presentation at the next meeting on their archive, once more information is known about the size.

- The following agencies are in contact with DSIG about their respective regional archives of AVHRR:

    o North American: NOAA and USGS. Presented at WGISS-55 and hopefully again at WGISS-58.

    o Canada: CCMEO. Hope to have them present at WGISS-58.

    o Hawaii: University of Hawaii. Tapes are currently being shipped from Hawaii to ESRIN.

    o Brazil: INPE

    o Argentina: CONAE

    o South Africa: SANSA

    o Central Africa (Kenya): ASI/UniRoma

    o China: CMA/NSMC.

    o India: ISRO. Presentation coming up.

    o Mongolia: ESA needs to get in contact.

    o Thailand: GISTDA. Don't preserve the archive of data.

    o Australia: CSIRO. Will continue working on the stitching of data.

− Still trying to identify other meteorological organisations who might hold data that could fill some of the current gaps.

− Several hundred heritage media have been collected, and slowly being transcribed into digital format. An ESA pilot activity has been completed with successful transcription of 2 different brands of optical disks originally held at ESA and University of Reading.

− A commercial company has been contracted to transcribe the full batch of available optical disks (around 522).

− In October 2023, ESA received the content of 15 transcribed optical disks. L1B products were successfully processed using the ESA processor.

− In February 2024, 50 media were successfully transcribed and data inventory under finalisation (media containing both AVHRR and Nimbus CZCS data).

WGISS-57-40	WGISS Chair to set up a meeting with Mirko Albani and USGS archive colleagues alongside WGISS-58 regarding the USGS AVHRR holdings (stitched data, documents, tools). Verify the USGS AVHRR holdings ahead of October.	**Due: WGISS-58**

8.8 - AVHRR data availability and access from Bhoonidhi

Sai Kalpana (ISRO) reported:

− ISRO acquires data from AVHRR instruments onboard NOAA 11-19, JPSS 1 and Suomi-NPP.

− This data has been catalogued and is available on Bhoonidhi platform.

− The AVHRR data of the NOAA series is available as open data on order. The product is disseminated through FTP. The JPSS and NPP Soumi products are available open data direct download and the product is disseminated through HTTP Access.

− The NOAA-AVHRR was used operationally in the Indian Space programme since 1989 under two important National Projects:

  ○ National Agricultural Drought Assessment and Monitoring System (NADAMS) (NDVI & LST)

  ○ Forecast Centre (MNCFC) under Min. of Agriculture & Farmers Welfare Potential Fishing Zones (PFZ) project (SST)

Discussion

– For NOAA-17 - 19, the raw datasets are available. ISRO will have to check whether the raw data is available for earlier missions, or if it is just the Level 1B data.

WGISS-57-41	ISRO to check the availability of AVHRR raw data, and prepare plots showing time period availability of AVHRR data from each mission.	**Due: ASAP**

8.9 - ESA FDR4AVHRR Project

Mirko Albani (ESA) reported:

– ESA is collaborating with UniBern to create a fundamental data record (FDR) from AVHRR data, and make this data accessible via the ESA dissemination service.

– The project kicked off in January 2024, and is expected to finish in December 2025.

– Tasks:

  ○ Extend the European data set with additional data from UniBern (2021 – 2024)

  ○ Integrate global data from South Africa, Argentina, etc. and from the 1km Land Project (1992 – 1999)

  ○ Develop a module "uncertainty" to be added to open source software PyGAC

  ○ Develop a module "orthorectification" to be added to open source software PyGAC

  ○ Improve the inter-satellite calibration for all channels

  ○ Reprocessing of all data (Europe, global) to Level 1C

  ○ Validation and quality assessment

  ○ Transfer all data to ESA for archiving and dissemination

– Co-operation with the EUMETSAT MetOp reprocessing team has started.

– The new FDR4AVHRR project will be a major contribution to ESA's Climate Change Initiative (CCI) and all studies related to Essential Climate Variables (ECVs). The dataset will have improved calibration, geocoding, validation, uncertainty, open source software (PyGAC), and extended 1km-AVHRR dataset (Europe and global).

Discussion

– Nitant Dube (ISRO, WGISS Vice-Chair) asked which ECVs the project is targeting. Mirko noted the project mostly focuses on snow cover, lake temperature, and perhaps others.

8.10 - DLR AVHRR data curation and reprocessing

Katrin Molch (DLR) reported:

- Timeline project was initiated 2013 for AVHRR LAC data consolidation, harmonisation, and geospatial product generation.

- AVHRR curation feasibility study was run with ESA from 2013-2015, to investigate the different approaches to a similar problem.
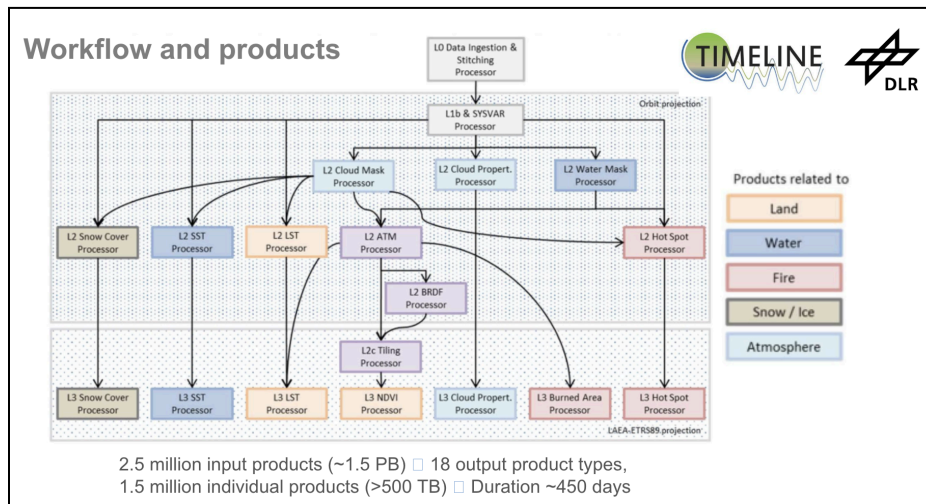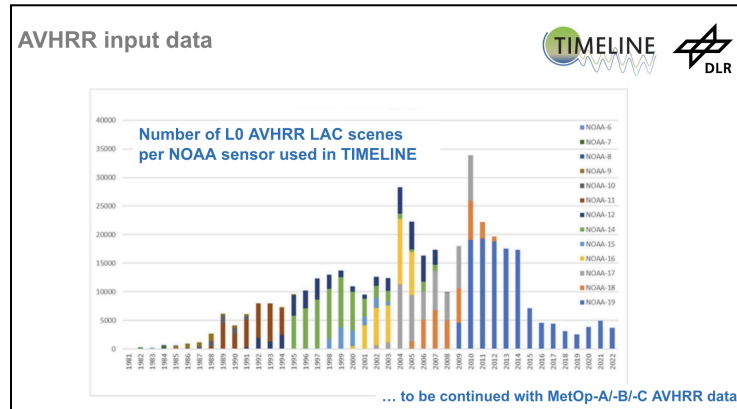
- First L2 and L3 products were published in 2022. Was a big effort to pull this together - with most of it happening in the first step of data harmonisation.

- The Timeline product suite includes NDVI, land surface temperature, SST, cloud properties, snow cover and burnt area. They are all available as daily, 10-day and monthly composites. Temporal resolution is from 1981-2022, with (in general) 1 km spatial resolution.

- Products are formatted in netCDF.

- L0, L1B and L2 data is for internal use. L3 projected composites will become free & open.

- STAC metadata will be generated for L3 products.

- CEOS-ARD self assessment ongoing for selected data, including L3 LST.

- Future access to Timeline L3 products will be via EOC-Geoservice for visualisation & download.

Discussion

- DLR has an interest in contributing to the international AVHRR effort. Would like to compare the L1B data generated by DLR to that by ESA.

8.5 - Data Stewardship and Maturity: WGISS DMSMM vs NCEI and WMO DSMM

Iolanda Maggio (ESA/Rhea) reported:

- Data Management and Stewardship Maturity Matrix (DMSMM) defines all activities needed to preserve and improve the information content, quality, accessibility, and usability of data and metadata.

- In creating DMSMM, DSIG reviewed many other maturity matrices and models relevant to this topic, including:

  ○ Capability Maturity Model Integration (CMMI)

  ○ Levels of Maturity of Digital Repositories

  ○ Climate Data Record Maturity Matrix (CDRMM)

  ○ Technology Readiness Levels (TRLs)

  ○ Scientific Readiness Levels (SRL)
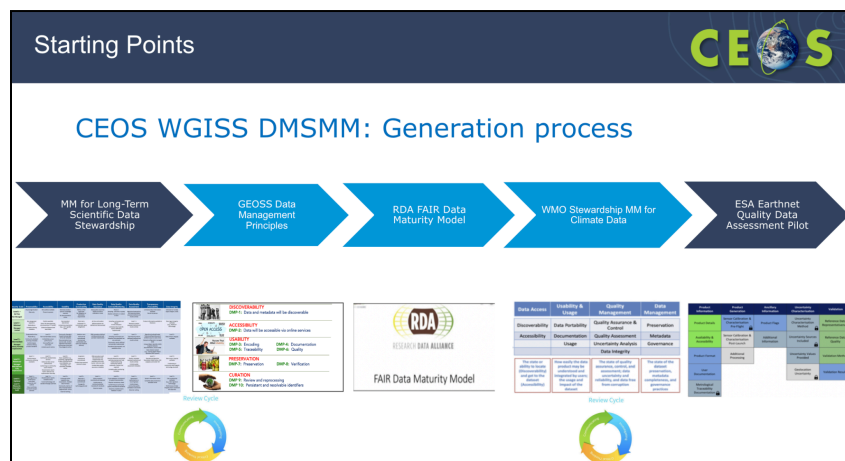
- 12 components in the DMSMM, in five areas: Discoverability, Accessibility, Usability, Preservation and Curation.

- There are four levels of maturity to measure against.

- DMSMM provides a way to measure the status of the agency data stewardship processes in place, and a way to plan goals of data stewardship processes and projects.

- DMSMM was compared against the NCEI/WMO article, CoreTrustSeal Trustworthy Data Repositories Requirements (CTDRR), Data Stewardship Maturity Matrix (DSMM) by NCEI/NOAA, and the FAIR guiding principles.

- CTDRR and DSMM are two of a handful of data stewardship maturity models used to ensure that best practices are followed for the acquisition, preservation, discovery, accessibility, interoperability, and reusability of data for the long-term. Both are based on the OAIS Reference Model. The main differences between these two approaches stem from their different perspectives.

- In contrast, the FAIR Principles provide high-level guidance for enabling that (meta)data are findable, accessible, interoperable, and reusable, and prompting data sharing in a machine-friendly environment. They have the most limited scope of the three approaches and tend to measure whether a Principle is satisfied or not. Unlike CTDRR and DSMM, they do not include maturity levels for measurement of individual FAIR Principles nor do they evaluate data quality practices explicitly.

- Strong correlations among these three assessment approaches are found in terms of data discovery, accessibility, interoperability, and usability as a result of their overlapping requirements for the stewardship of data objects.
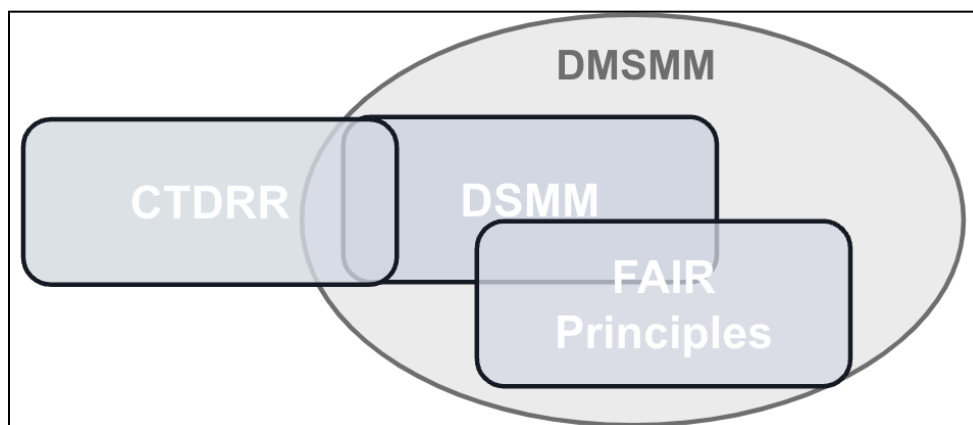
- WGISS DMSMM is now mature and used at ESA.

- There is an open action for WGISS DMSMM to be presented in a webinar organised jointly with WGCapD. Not sure whether this should be continued.

Discussion

− Nitant Dube (ISRO, WGISS Vice-Chair) will contact DSIG for support on developing an Interoperability Maturity Matrix.

− Tom Sohre (USGS, WGISS Chair) is not sure on the role of WGCapD for the webinar.

− There is a self-assessment document already, and the hope is to get a lot of agencies assessing against this. All WGISS members should take a look at this.

− Iolanda has created a flow chart to help people understand how to use the matrix.

− The white paper describes the content, and explains the components and how to reach the levels.

− Suggest next WGISS have a presentation from NOAA. They have developed a tool which automates a score. DSIG could consider developing something similar for DMSMM.

WGISS-57-42	DSIG to consider an annex to the DMSMM white paper containing a guide on how to use the maturity matrix.	**Due: WGISS-58**
WGISS-57-43	DSIG to invite NOAA to present on their maturity matrix COMET tool at WGISS-58, and consider whether WGISS should develop a similar tool.	**Due: WGISS-58**

Day 4: Thursday 7th March, 2024

Session 9: Collection Management

9.1 - Session introduction

Mirko Albani (ESA) and Tom Sohre (USGS, WGISS Chair) reported:

− Goal is to look at a future white paper to look at any best practices regarding collection management, and will start by sharing experiences today.

− Data collection management challenges include integrity, authenticity, replica management (in the cloud and/or archived), reproducibility and citation.

− The objective of this session is to discuss data collection management practices, starting from approaches in place at WGISS Agencies.

− Themes to include in the discussion are:

  ○ Common understanding and definition of "Data Collection" concept;

  ○ Preservation of collections;

  ○ Reproducibility of collections (algorithm availability);

  ○ Collections data citation;

  ○ Cross-collection validation (for example, NASA's "golden month" concept);

  ○ Approach to transitioning from one collection to another (time overlap, communication, versioning, etc.);

  ○ General Interoperability and governance approaches.

Discussion

− Richard Moreno (CNES) noted that for Copernicus, often only 2-3 months of data are stored, and it is hence not a complete collection. There are also issues with gridding and federation.

− Steve Covington (USGS) highlighted the issues relating to regional data hubs and partial replications. For example, the data ingested into Digital Earth Africa, was that data transferred correctly?

9.2 - ESA collection management practices

Ferran Gascon (ESA) reported:

- Collections management enables collections (including metadata) to be properly handled and documented, and to become a resource for the user community.

- The benefits of collections include:

  ○ Consistent data accessibility, usability, and long-term preservation.

  ○ FAIR (Findable, Accessible, Interoperable, and Reproducible) scientific practices.

 ○ Facilitate the reanalysis of satellite measurements.

- Collections are often generated through the reprocessing of the archive and therefore also referred to as "reprocessed datasets".
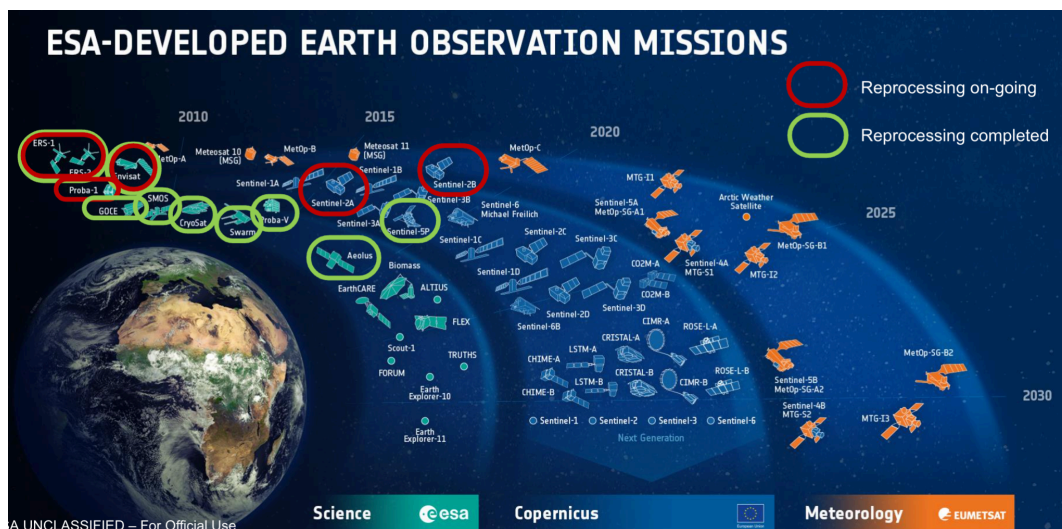
- The ESA collections include ERS, Envisat, Proba-1, GOCE, SMOS, CryoSat, Proba-V, Aeolus, Copernicus Sentinel-2 and Copernicus Sentinel-5P

- Copernicus Sentinel-2 Collection 1 provides consistent time series with a uniform processing baseline. The collection contains both Level-1C (top-of-atmosphere) and Level-2A (bottom-of-atmosphere) products. Ingestion of the Sentinel-2 Collection 1 data into the Copernicus Data Space Ecosystem will be completed in Q2 2024.

- The 4th reprocessing of data from the AATSR instrument flown on Envisat has been released to users. This covers the full mission archive from May 2022 to April 2012, in Level 1B data.

- Fundamental Data Records for Altimetry (FDR4ALT) is an ESA project aiming at generating innovative Earth system data records. ESA will reprocess ERS and Envisat Altimetry data for this project.
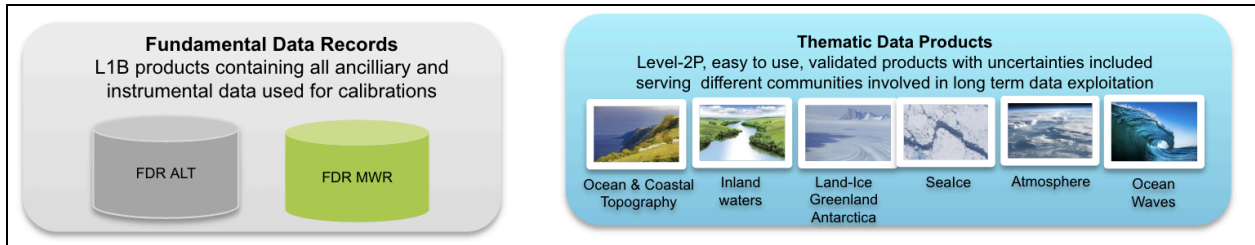
- Scientific applications, such as those for climate change monitoring, have underlined the importance of collections management.

- The development of new "down-stream" applications and services using space-borne data have increased the need for collections management.

- There are heterogeneous approaches for collections management, both within ESA and across agencies.

Discussion

- Tom Sohre (USGS, WGISS Chair) asked how ESA handles previous collections. When a collection is removed, what is the protocol for keeping it or not? Is there a way to gain access to a previous collection data?

- Ferran noted ESA does not have a harmonised approach across the agency for handling past collections. For Sentinel-2, initial data will not be kept, but it remains to be seen as to whether Collection 1 is kept when Collection 2 is developed, noting the large cost associated with storing the archives. It may be that just the latest collection is kept.

- Tom noted that USGS have similar concerns, and take a similar approach.

- Mirko Albani (ESA) recognised it is decided case by case, and depends on the size of the archive. For heritage missions, the previous collection is also kept, but on tapes, so it doesn't cost so much.

- Tom asked who could access previous collections, if it was decided they should be kept? What criteria is needed to access? Mirko would have to check about how many people ask.

- Steve Covington (USGS) noted that as USGS prepares Collection 3, they want to work with other Agencies to get closer towards interoperability.

- Does ESA have any guiding principles on inconsistencies in upgrading collections? USGS is aware of some issues in Landsat Collection 2, but is holding off until Collection 3 to make the updates to ensure consistency.

- Ferran recognised this is one of the main open points, and there is no common answer across ESA. For Sentinel-2, there are two options: freeze the baseline, or make a small upgrade, ensuring that as long as the product is improved with respect to the collection and there is no negative impact for users. At the same time, ESA does not want to disturb the work of users. However, they recognise that one cannot stop the evolution of the products waiting for the next evolution of the collections. Adding new layers makes it a bit more tricky when it comes to data quality.

- Steve recognised this is a place where WGISS can offer help and guidance for all CEOS Agencies. Help all come to a conclusion regarding the conflicting priorities when it comes to collection management.

- Ferran would like to have some best practices, which should be developed in consultation with users.

- Nitant Dube (ISRO, WGISS Vice-Chair) asked about the management of collections hosted in the Copernicus regional data hubs, where data is pushed onto various clouds. When a user finds and accesses a product via CMR and FedEO, how do they get information about collection availability? How does a user come to be aware that data is available on regional hubs as well?

- ESA communicates to users via the regular communication channels for each mission, and the news feeds for the various collections. If there is an update, the catalog is updated, and it is harvested immediately by FedEO.

- Nitant recognised there are issues to address with collection replication. If there is an update made by ESA, connections should be made with those who hold copies of the data.

- Alex Leith (SEO/Auspatious) noted the example of Microsoft Planetary Computer, who don't have a list of recently updated datasets. There is no way to find out whether Microsoft has the most recent copy.

- For the big reference collections, such as Landsat and Sentinel-2, it would be good if they had some sort of feed to indicate which scenes were reprocessed and updated.

- Ferran suggested that one important point to include in the best practices would be the format and content of the metadata, to assist access via APIs. For Sentinel-2, there isn't any reference to specific collections in the metadata.

- Alex noted that Element84's copy of Sentinel-2 calls it Collection 1 specifically.

9.3 - USGS collection management practices

Tom Maiersperger (USGS) reported:

− USGS have a number of data management practices which help guide their collection management.
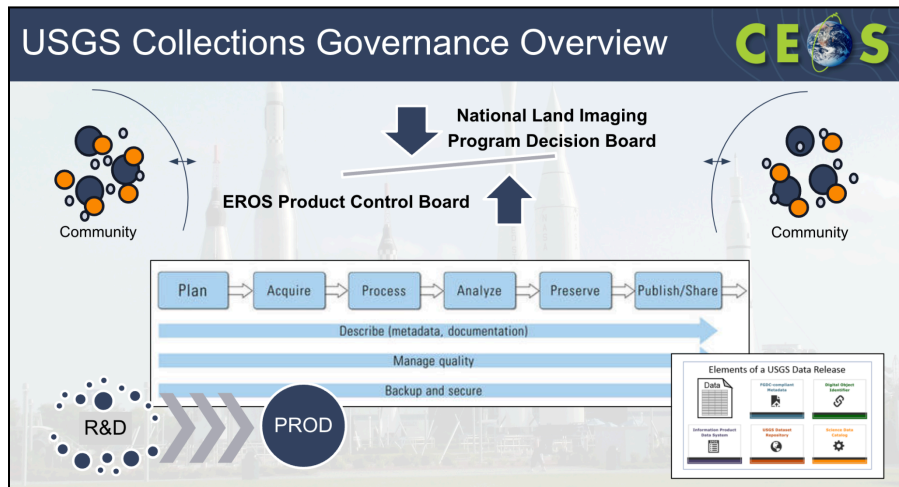
− Defining the terms is key. The terms USGS use are as below:

Landsat Levels of Organization		Macroecology	
Collection	Landsat Collection 2	Community	
Dataset	Collection 2 Level-1 Landsat 8-9 OLI/TIRS	Population	
Product	LC09_L1TP_042026_20220809_20230403_02_T1	Organism	
File	LC09_L1TP_042026_20220809_20230403_02_T1_B1.TIF	Organ	

− Datasets have DOIs, and data format control books. Datasets are the fundamental published work.

− Datasets are organised by like-sensors.

− US ARD products are available for the continental US, Alaska and Hawaii, with equal area gridded products available for Level 1 and 2.

− Minor revisions to products in a collection can be expected, but are infrequent and small in magnitude. To maintain the highest quality and consistency within the collection specification, this is typically in the event of errors or anomalies.

- Revised products supersede lower quality products and are identifiable by processing date.

- Changes to lower-level products propagate to higher-level products in the processing lineage.

- The nature and magnitude of minor reprocessing events are communicated to end users via a [news feed](#).

- The only thing that lets a user know that it is a later reprocessing is a new date. There is no versioning numbering, but this is something under consideration for Collection 3.

- Changes allowed without a full processing include unavoidable changes in forward production, dataset-appending changes and collection-appending changes.

- Major changes to collection specifications are saved for a full reprocessing of the historic record in the next collection.

- Collection 2 refers to the "2nd major version of improvements across the datasets"

- Once a new collection is processed and released, the prior collection is superseded. USGS follows a few steps for this process:

  - Deprecation and User Transition: Cease forward processing of new data acquisitions into the superseded collection after ~ 1 year in parallel with new collection, then retain accessibility of the superseded collection for an additional ~ 1 year to allow user transition to improved data.

  - Decommissioning: Make the superseded collection inaccessible; Retain access to tombstone digital object identifier pages and documentation related to datasets (DFCBs, ADDs); Publish a legacy sample of superseded products for future data forensics and intercomparisons.

  - Preservation: Retain and back up the lowest level of raw instrument data or other baseline data in perpetuity. Superseded collections may be deleted.

- Users are kept informed of the changes throughout the process.

Discussion

- Richard Moreno (CNES) noted that for the Copernicus cloud archive, there is a proposal under development for services to process data on demand. This is a way to keep the products available, and allows users to adapt to the changes in processing

- Tom recognised USGS don't have that available at the moment, but there are some studies looking at virtual collections. A lot of the infrastructure is tied to the

on-premises data storage, and USGS is going through a process at the moment to redefine the processing chain to be cloud-based. Would like to have it available on the cloud in the future, but there are lots of legacy elements to deal with first.

−   Steve Covington (USGS) recognised the historical nature of the data provided should be considered. There are risks in allowing people to process their own data, as the government has a responsibility to provide accurate datasets. This includes careful and deliberate choices in what makes the data quality appropriate.

−   Tom noted USGS have only had a handful of users wanting access to Collection 1.

−   Richard suggested there is no responsibility for the data provider to ensure the data user's product is the same as the one provided. Agencies should allow users some capability to process their own data.

9.4 - NASA collection management practices

Doug Newman (NASA) reported:

−   NASA's definition of collection is different to USGS's. As per the Data Product Development Guide for Data Producers:

   ○   Data Collection - A major release of a data product, or of a set of closely related data products, which can be followed by minor releases within the same collection.

   ○   Data Product - A set of data files that can contain multiple parameters, and that compose a logically meaningful group of related data.

   ○   Dataset - A broadly used term that can be used to describe any set of data. The official term "HDF5 dataset" describes a data array in an HDF5 file (equivalent to a NetCDF-4 variable in a NetCDF-4 file or an HDF-EOS field in an HDF-EOS file). At the opposite extreme, an entire data collection can also be referred to as a dataset

−   NASA aims to ensure data collections are secure, accurate, available and usable.

−   A data governance board creates working groups to develop standards and guidelines for NASA's Earth Science Data System (ESDS) and Earth Science Division (ESD). One such group is the Data Producers Working Group who are responsible for the Data Product Development Guide.

−   Approved standards are developed and endorsed through the ESDIS Standards Coordination Office (ESCO).

−   Integrity of data is of utmost importance. All data in transit will be encrypted using full end-to-end TLS encryption by September 15, 2024.

- The accuracy of data is controlled by interface control documents for mission-level data products and Earthdata Pub for smaller products.
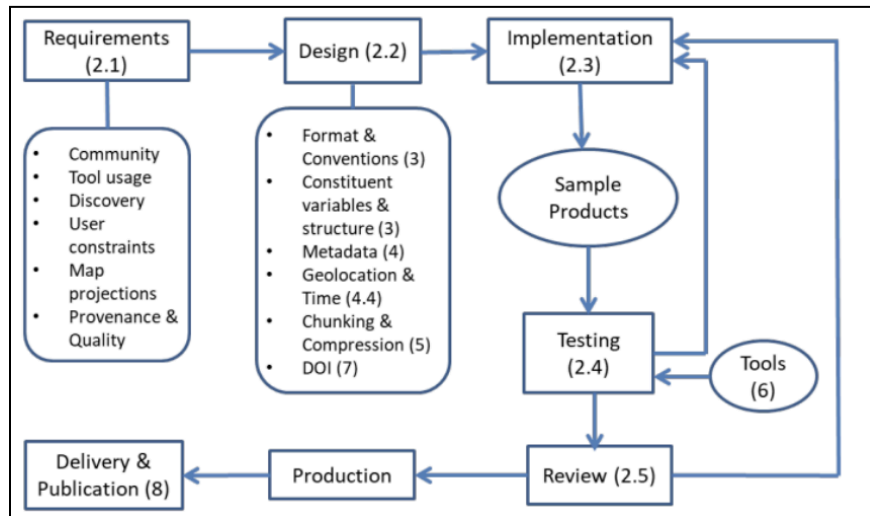

- The archival model uses a three copy approach:

  ○ Public, accessible copy of all data in AWS S3

  ○ Non-public, back up copy of all data in AWS S3 (separated by AWS account) - NASA Earthdata ORCA

  ○ Non-public, tape copy of all level zero data on-premises - NASA Earthdata LZARDS

- NASA has a number of collection services to help add value to the data, including visualisation services.

- Data governance acts as a gate between steps in the data lifecycle, and now also encompasses getting the data fit for purpose in cloud ecosystems

- NASA has a large community of experts for evaluation and associated collections in CMR to ease discovery of services.

- The problems identified include:

  ○ Consensus on data takes time, and there is an explosion and contention of cloud-optimised data formats.

  ○ The trade off between speed and storage costs of intermediary formats

  ○ As services become more sophisticated so does their metadata

  ○ Planning and direction is needed to avoid a 'one service/tool per collection' model

9.5 - GA collection management practices

Simon Oliver (GA) reported:

−   The CEOS Data Stewardship Glossary defines a collection as "The ensemble of some products or auxiliary data having a common focus or theme or purpose (e.g., collection of land photos)."

−   From GA's perspective, the attributes of a collection are:

  ○   Consistent base

  ○   Contains multiple datasets derived from a common root

  ○   Some change is allowed, within tolerance

  ○   Multiple factors can drive decisions regarding management of collections
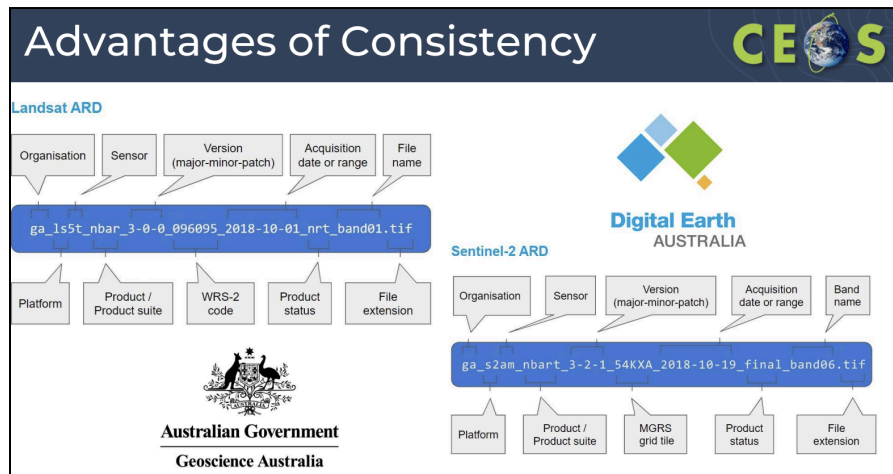
−   Consistency between the collections are maintained through a common code base, with consistent approaches to ARD generation.
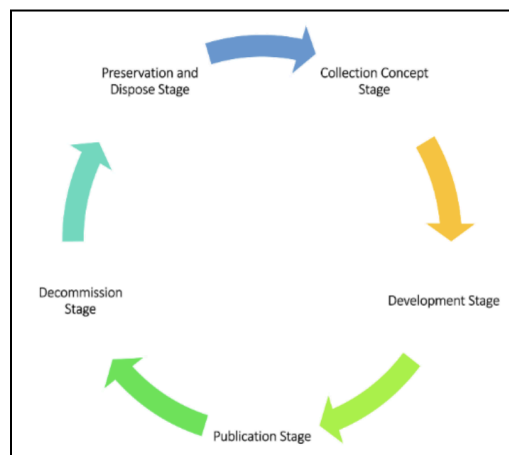
– The collection governance in GA is guided by the National Archives Act of 1983, the GA information and records management team, the Digital Earth Australia Management Committee and the Digital Earth Australia Control Board.

– A collection update can include software updates, source or auxiliary data reprocessing or upgrades, metadata fixes and remedial reprocessing.

– A collection upgrade would include algorithm enhancement or baseline/input data upgrade.
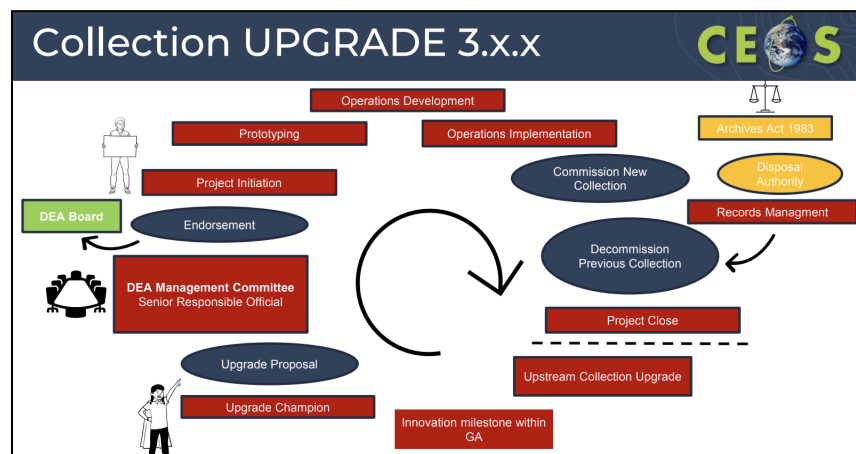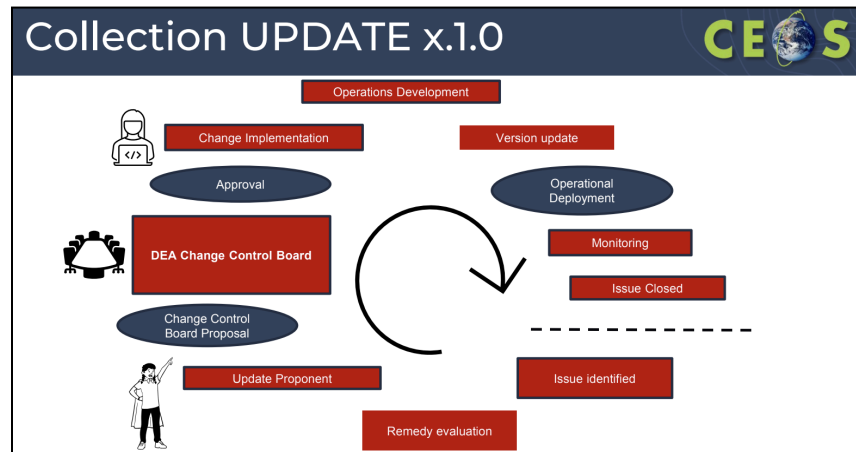
– Updates to source datasets are staged and not automatic

– USGS Landsat 9 reprocessing does not automatically trigger updates in the GA Collection.

– Changes to data or systems with a potential to impact the consistency of the Collection must be approved by the Change Control Board.

– Upgrades to Collections are approved via the Digital Earth Australia Management Committee.

− Collection Disposal is conducted in line with Archive Act responsibilities through consultation with Records Management and National Archives of Australia.

− All changes likely to impact users must be communicated.

− Only a single instantiation of a dataset (near-real time, interim or final) is available to a user at any given instant. Dataset maturity may be specified in API calls. The availability of near-real time data via a single stream was requested by users.

− GA supports WGISS developing a set of common collection management principles.

Session 10: Biodiversity

10.1 - Context: Biodiversity in CEOS

Gary Geller (NASA, EETT Co-lead) reported:

− Biodiversity describes the variety of life on Earth. It covers genes, species and ecosystems. Ecosystems are the key area where satellite Earth observation data can support.

− The services provided by ecosystems are important for human survival.

− The world is tackling a global biodiversity crisis.

− CEOS started working on biodiversity in 2011, but this really kicked up in 2021. The Ecosystem Extent Task Team (EETT) was established at the 2022 CEOS Plenary, and the white paper was presented at the 2023 CEOS Plenary. In 2024, EETT is managing three demonstrators.

− Remote sensing can observe:

○ Environmental characteristics (temperature, precipitation, topography)

○ Optical characteristics of vegetation (Greenness, leaf & canopy "traits" (water content, structural compounds…))

○ Structural characteristics of ecosystems (height, profiles) via radar and lidar sensors

− It is important to combine data from various sources of data.

− There are better capabilities now for enhanced sensors, more computing power, and new technologies.
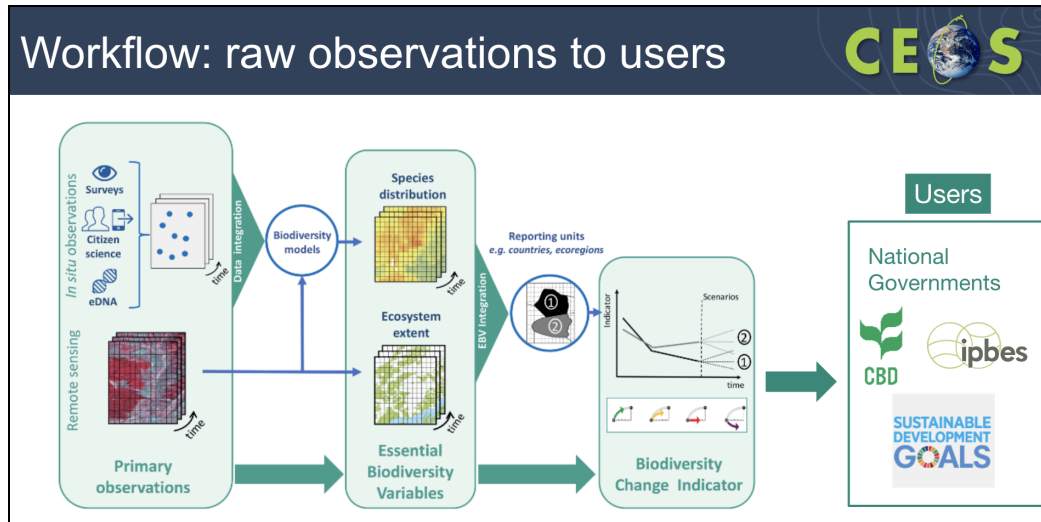
- Some opportunities for biodiversity and remote sensing include:

    ○ Combining data from different types of sensors

    ○ Use of hyperspectral data

    ○ Artificial intelligence and machine learning

    ○ Data cubes

    ○ Global monitoring
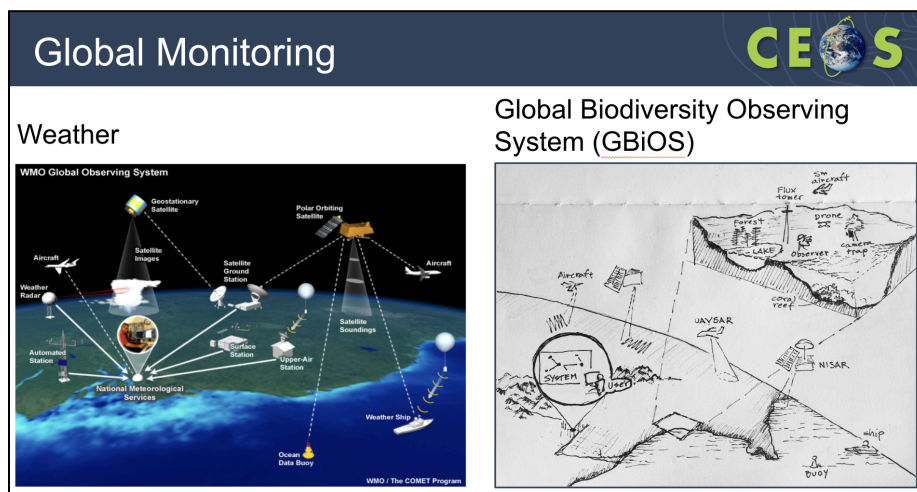
- Interoperability of remote sensing datasets is key for biodiversity work.

- Challenges faced for remote sensing and biodiversity include:

    ○ Combining data from different types of sensors

    ○ Operationalizing workflows to create value-added products

- ○ EO data usability & technical capacity of users

- ○ Standards: Interoperability & comparability

- ○ Global monitoring

10.2 - Technical Challenges - 2023 EETT White Paper

Roger Sayre (USGS, EETT Co-lead) reported:

- − Roger is a spatial ecologist, working on ecosystem mapping. Not specifically an expert in remote sensing data.

- − The genetic aspect of biodiversity is not remote-sensing detectable at this time.

- − Detection of fauna is not practical with remote sensing technologies available at this time, while flora is increasingly detectable with hyperspectral imagery.

- − Ecosystems occupy space in three kinds of spatial configurations: small patch, large patch, and matrix-forming. Many terrestrial, freshwater, and coastal (e.g. shallow benthics) ecosystems are detectable via satellite imagery, and mapping Ecosystem Extent is an application area within Biodiversity.

- − There are three fundamental characteristics of ecosystems:

  - ○ Ecosystem Structure: Practical approach to mapping ecosystem extent. Map ecosystems by mapping and combining their structural components (vegetation assemblages, climate settings, terrain setting). Image-based approach with modelling.

  - ○ Ecosystem Function (Processes): Some aspects of ecosystem function are mappable (e.g. productivity, flux of water) but many are not (nutrient cycling, competition and predation, etc.). There are image-based approaches available for some functions.

  - ○ Ecosystem Composition: Addresses biotic composition, largely biogeographic in nature, captures differences in species distributions, based on field/expert knowledge. There is not an image-based approach for composition.
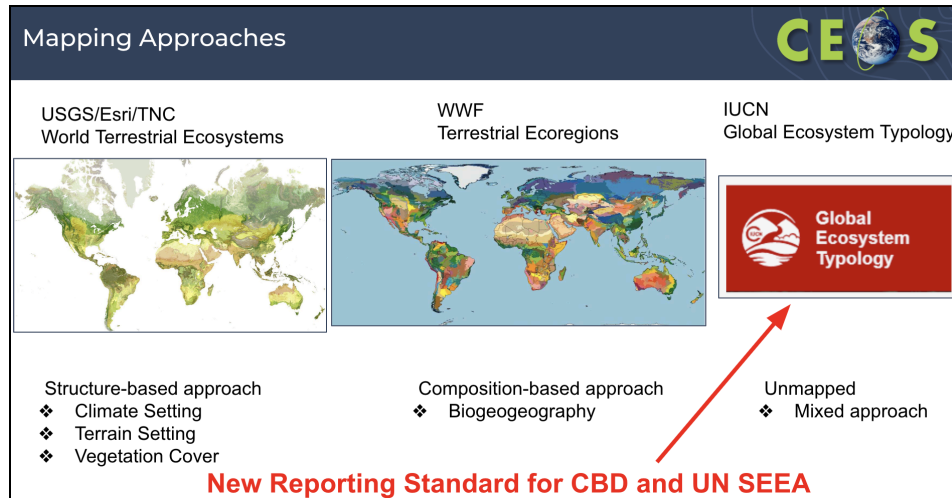
- The GEO Global Ecosystem Atlas activity focuses on amassing all good ecosystem maps at country, regional and global scales, and making them publicly and easily accessible.

- Global Ecosystem Typology (GET) maps do not yet exist for most countries. National maps of GET ecosystems will be produced - this is one of the main priorities of the GEO Global Ecosystems Atlas.

- All three of the EETT demonstrators are using data cube structures. Using multiple data cubes, with both x, y, z dimension for 3D spatial cubes, as well as x, y, t for temporal evolutions.

- Data cubes are not well adopted within the biodiversity community, but uptake is on the rise.

- Data cubes work when data structure is raster. The biodiversity community often uses vectors and polygons instead, which does not easily slot into data cubes.

- Data formats used include:

  ○ Vector data: shapefiles, GeoJSON, kml/kmz

  ○ Raster data: netCDF, GeoTiff

  ○ Tabular and point data: CSV

  ○ Software: ArcGIS, QGIS, Google Earth Engine

- Open source software is very important for the biodiversity community.

- Cloud availability of data helps ease the weight on having data locally.

- One standard in the ecosystem mapping community is the IUCN GET. Standard reporting unit for the UN Convention on Biological Diversity (CBD) is the functional groups.
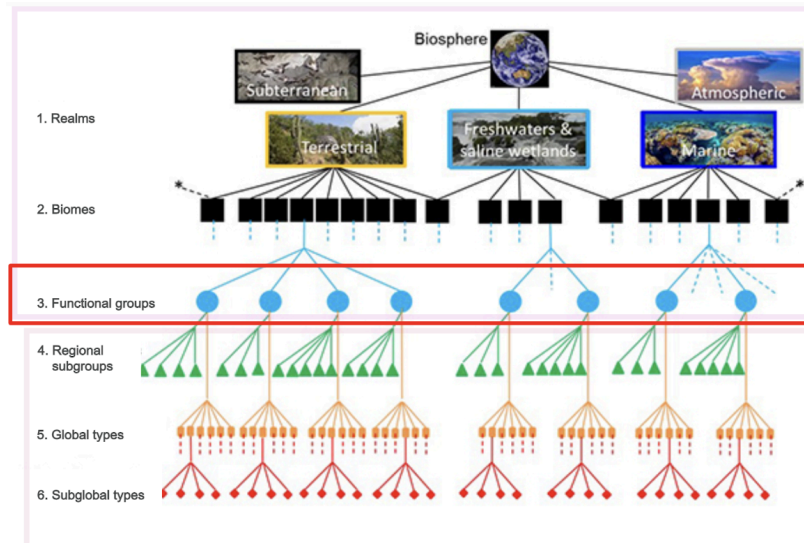
- All five factors of the interoperability framework are important for biodiversity work.

- Wall-to-wall ecosystem mapping by image classification AI/ML is 'old news'. 'New' AI (Geo AI) has at least two potentially powerful applications:

  - Object-based feature extraction (mangroves, burned structures, roads/trails, etc.)

  - Crosswalking between existing maps to the IUCN GET using labels, attributes, etc.

- Challenges for WGISS to consider include:

  - Limited availability of value-added products

  - Combining data from different types of sensors

  - EO data accessibility, usability and technical capacity of users

  - Ecosystem condition - mixed with Ecosystem Extent

  - Reference data for training and validation

  - Scale

10.3 - Hudson Bay Lowlands Demonstrator

Jason Duffe (ECCC) reported:

- Jason is a biologist, but manages a team of geographers and remote sensing experts, and is working on linking the two communities.

- Hudson Bay Lowlands (HBL) covers a 320,000 km^2 area in central-eastern Canada. HBL is a peatland, and emits megatons of carbon.

- The objective is to assess biodiversity and carbon values, in particular for peatland, wetland, coastal and marine ecosystems in the Hudson Bay-James Bay Lowlands, and weave western science with indigenous knowledge to inform the range of conservation efforts in the region.

- Aiming to draw lines around different ecosystem types, describing contiguous groups of plants, etc. This helps inform the biodiversity and climate change reporting.
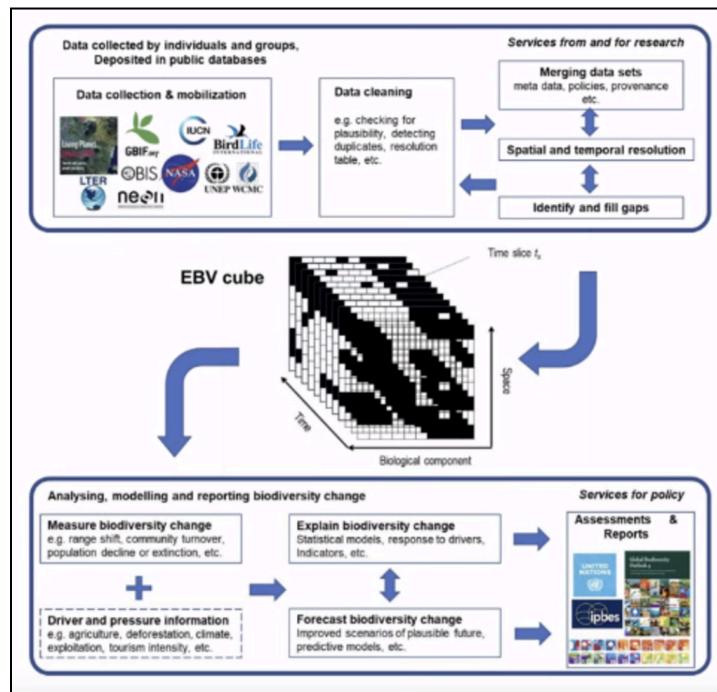
- ECCC has been using data cubes since 2016, with funding from CSA to demonstrate a Canadian implementation of the Open Data Cube.

- The overarching purpose for the demonstrator is to convey how satellite remote sensing is essential for mapping and monitoring ecosystems.

- The demonstrator will use Machine Learning Classifiers to create a detailed ecosystem map of Wapusk National Park. The work will use advanced AI algorithms and other methods to combine data from several satellite sensors, in situ biodiversity data and data from other sources.

- It uses the CEOS Analytics Laboratory developed by the SEO. Will use Landsat and Sentinel data, and are hoping to obtain hyperspectral data from EnMAP, and radar data from ALOS (L-band) and TanDEM-X (X-band).

- The below image was created ~15 years ago with Landsat data. The goal is to do this level of mapping across the whole ecozone.
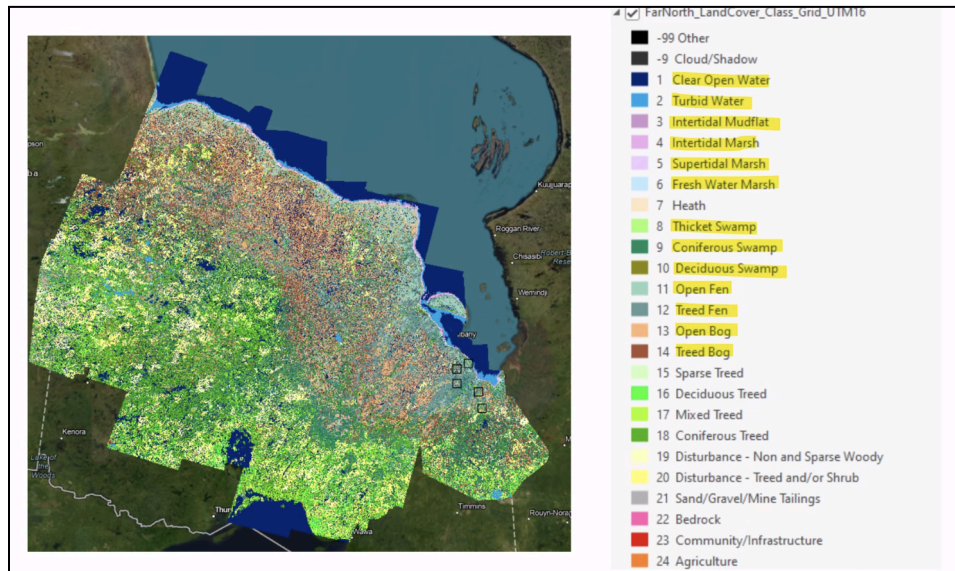
- Challenges and opportunities include:

  ○ New sensor platforms - including NISAR launching this year.

  ○ Use of hyperspectral data in the data cubes

  ○ Combining hyperspectral and SAR data using neural networks for ecosystem classification.
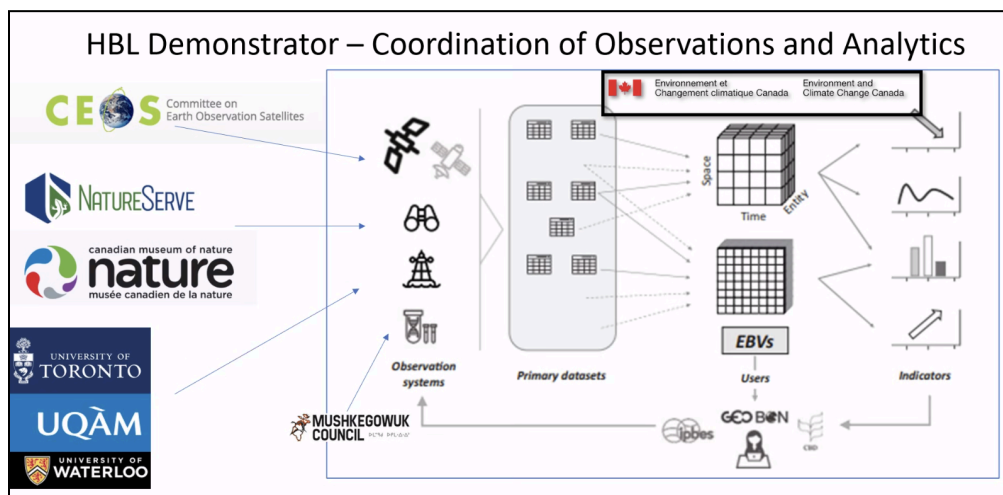
- Carleton University is looking at developing approaches for interoperable management of biodiversity data from different knowledge systems.

10.4 - Discussion: How can WGISS support Biodiversity work in CEOS?

− Tom Sohre (USGS, WGISS Chair) is unsure where the WGISS role is here, and there might be more to learn from the biodiversity community. Great to see them working with the SEO and CSIRO on the data cube.

− There is an opportunity to take lessons learned from biodiversity activities to support the interoperability framework.

− AI topics might be another area to build on.

− Steve Covington (USGS) noted the use of hyperspectral data, and bringing in commercial data sources. This is a great example of multi-source, multimodal applications of EO data. Are CEOS Agencies putting in the place the right systems to make use of the disparate data sources?

− Roger Sayre (USGS, EETT Co-Lead) recognised hyperspectral is very useful for delineating different vegetation types. The EETT work will likely produce some lessons learned about how to bring hyperspectral datasets in with other datasets.

− Gary Geller (NASA, EETT Co-Lead) asked whether WGISS has experience with other disciplines in combining sensors to create products, as this is something the biodiversity community is struggling with. Additionally, when generating workflows and ensuring continuity along the entire processing workflows, there are so many places to go wrong. Data cubes can help here, and hence why all three demonstrators are using data cubes. However, there are challenges with loading data up and using it locally.

− Matt Paget (CSIRO) noted the CEOS-COAST group may be in a similar work flow evolution space. They are also working with the CEOS Analytics Lab.

− Tom thanked the EETT Co-Leads and Roger for introducing the biodiversity topic to WGISS.

WGISS-57-44	WGISS Exec to discuss how WGISS might be involved in the CEOS biodiversity activities, responding to the challenges presented at WGISS-57.	**Due: Q2 2024**

Session 9: Collection Management (cont.)

9.6 - Collection Management Discussion

− Steve Covington (USGS) recognised that if there was more global compatibility between collections, users would be able to easily search across different collections.

- Peter Strobl (EC-JRC) is working on product level definitions, which is another key aspect of collection management.

- Tom Maiersperger (USGS) recognised two aspects of collections management: content and structure. Data providers often talk about product levels, while users are thinking more about the parameter aspect.

- Steve asked whether it would be a tangible goal to come up with a more user centric view of collections across the agencies?

- Patrick Quinn (NASA) questioned how this would differ from the various ARD efforts.

- Steve noted the USGS Landsat product isn't called "Surface Reflectance", it is simply compliant with the CEOS-ARD surface reflectance specifications.

- If USGS and GA both produce Landsat products compliant with the CEOS-ARD Surface Reflectance specifications, but the USGS product doesn't have BRDF while the GA product does, they both are CEOS-ARD compliant, but non-compatible.

- On the topic of collection version control, WGISS should aim to coordinate across the agencies to phase the lifecycle of collections to improve user experience. This includes conventions on naming and differences between major version, minor versions and bug fixes.

- The terminology discussion comes up in almost every WGISS topic, and this should be corrected first, but it is unclear if WGISS Agencies can come to an agreement. Is there a way to provide a mapping between different terms?

- Alex Leith (SEO/Auspatious) noted STAC's convention is pretty simple, and perhaps CEOS Agencies can try and align to STAC definitions.

- Tom Sohre (USGS, WGISS Chair) raised concerns about aligning to a technology which may not exist in 10 years.

- Is there a benefit of trying to normalise everything from nomenclature to product equivalencies?

- Patrick recognised harmonising nomenclature might be more beneficial than other parts. It would be great to see a list to see of topics where the most impact can be made, e.g. version control, language.

- Mirko Albani (ESA) noted the GA/USGS document provides some good starting points, and has some answers already to some of the questions raised.

WGISS-57-45	GA and USGS to share their work on collection	Due: ASAP

| | management to the broader group for review via GitHub (coordinate with Libby). DSIG will organise a meeting in May 2024 to discuss the work done by GA and USGS, and the way forward on a WGISS white paper. | |

Chat Messages

Alek Cesarz (CDSE): CDSE includes OnDemand Production service which allows creation of products using different (historical or future) baselines. This in part solves the issue of frozen baseline for a collection - the mainstream collection follows a single baseline "family", but when a new baseline with breaking change is introduced it would be possible to make it available to some users so they can produce improved products for their AOI/TOI.

Alek Cesarz (CDSE): Is the relation between old (obsolete) and new product stored in the metadata? Or detectable in API?

Tom Maiersperger (USGS): Yes, collection number is contained in both the dataset-level and product-level metadata. An individual product that has been superseded within an existing collection is simply "revised and replaced" today. Processing date becomes the discriminator. See https://www.usgs.gov/landsat-missions/landsat-collection-2-reprocessing-events

Patrick Quinn (NASA): Tom mentioned changing of bit packing being a breaking change. I'm curious about this and whether we think things like compression and chunking decisions constitute changes to the data.

Patrick Quinn (NASA): Users should be able to reproduce what we do. They don't have the canonical copy, and they certainly can mess it up, but we don't hold a monopoly on correct processing.

Tom Sohre (USGS, WGISS Chair): I think the point is that we need to be able to differentiate a product created by various processing streams (one being the authoritative agency stream and others potentially being "open science"). How does NASA handle the transition between collections (or versions)? Timing of overlap? Maintenance (or not) of retired collections (versions)?

Patrick Quinn (NASA): It varies between data providers and datasets, but in general when a new version is created, we will keep the prior version for some period of time on the order of years, and deprecate / delete older versions.

Tom Sohre (USGS, WGISS Chair): I think the transition between collections (including overlap of availability, communications, and handling of dated/retired data) being an important discussion area for consistency (user perspective).

Patrick Quinn (NASA): Much of the variation we have is user-driven, though. What are the community/user expectations? How popular is the dataset? How much are older versions being used or referenced? Communication is definitely key, though, and it can be hard, especially for programmatic access or data cited in publications

Alex Leith (SEO/Auspatious): You know, it might be good to agree on and move towards a common definition of dataset/product/collection/granule/scene and so on! Sounds like Collection is pretty much agreed, but the tier or tiers below that could do with a common terminology.

Session 11: Closing

11.1 - Review of actions and decisions

Libby Rose (WGISS Secretariat) reviewed the preliminary Actions and Decisions record.

11.2 - WGISS-58 Planning

Tom Sohre (WGISS Chair) reported:

- WGISS-58 will be held October 15-18 at USGS EROS in Sioux Falls, South Dakota, USA.

- This will include a joint meeting with WGCV.

- The Chair team will scope out the possibility of an industry engagement workshop.

- WGISS is also looking for host Agency for WGISS-59 (~April/May 2025) and WGISS-60 (~Sept/Oct 2025)

- Tom would like the fall meeting to focus on Work Planning for the following year, and the spring meeting to discuss the status of that Work Plan.

11.3 - Closing Remarks

Tom Sohre (WGISS Chair) reported:

- WGISS-57 had around 50 participants during the week, both in-person and virtual.

- Great discussions across the whole week on a variety of topics.

- Reminded participants to encourage others from their agency to get involved in the Interest Groups.

- Nominations for WGISS Vice-Chair for 2026-2027 (and subsequently WGISS Chair for 2028-2029) are welcomed.

Appendix A: List of Participants

In-person participants	
Affiliation	**Name**
CNES	Richard Moreno
CSIRO	Matt Paget
ESA	Mirko Albani
ESA/Rhea	Iolanda Maggio
ESA/Solenix	Filippo Marchesi
ESA/Spacebel	Yves Coene
GA	Claire Krause
Geoscience Australia	Simon Oliver
ISO/TC-211	Liping Di
ISO/TC-211	Liying Guo
ISRO	Nitant Dube
JAXA	Yousuke Ikehata
NASA	Doug Newman
NASA	Patrick Quinn
SEO/Auspatious	Alex Leith
SEO/Symbios	George Dyke
USGS	Thomas Maiersperger
USGS	Tom Sohre
USGS/Aerospace	Steven Covington
VNSC	Nguyen Tien Cong
WGISS Secretariat	Libby Rose

Virtual Participants	
Affiliation	**Name**
AEM	Adrian Guzman
CEOS Executive Office	Irena Drakopoulou
CEOS Executive Officer	Steven Ramage
Copernicus Data Space Ecosystem	Alek Cesarz
CSA	Carmen Marcu
CSA	Lucie Viciano
DLR	Felix Feckler
DLR	Katrin Molch
DLR	Mario Winkler
EC-JRC	Peter Strobl
ECCC	Jason Duffe
Element84	Matt Hanson
ESA	Damiano Guerrucci
ESA	Ferran Gascon
ESA/Freelance	Matthias Mohr
Eurac Research	Alexander Jacob
Geoscience Australia	Alison Rose
GISTDA	Prayot Puangjaktha
GISTDA	Saowanee Srichai
ISRO	Ashutosh Gupta
ISRO/NRSC	Sai Kalpana
LSI-VC Sec	Matt Steventon
NASA	Dana Ostrenga
NASA	Minnie Wong
NASA / EETT	Gary Geller
NASA/KBR	Michael Morahan
NASA/SEO	David Borges
NOAA	Kenneth Casey
NOAA/CEOS COAST	Prasanjit Dash
SEO/AMA	Brian Terry
SEO/AMA	Josh Baptist
USGS	Roger Sayre

Appendix B: Decisions

Decision 01	The Archive Technologies White Paper was endorsed by WGISS-57.

Appendix C: Actions

WGISS-57-01	Libby to work with Exec to find a place on the website for external documents as well as a process to propose new linked documents and review existing links for continuing relevance.	**Due: ASAP**
WGISS-57-02	Mirko, Richard, and Dave Borges (SEO) to coordinate with GEO POC (Paola De Salvo) to provide GEO data topics (incl. GIDTT status) presentation at WGISS-58.	**Due: WGISS-58**
WGISS-57-03	WGISS to demonstrate the use of a moderated repo approach (GitHub) for Best Practices (or other document) development/feedback. Libby to coordinate with SEO to identify processes (and training) for WGISS staff on CEOS repository. Libby/Doug to coordinate with SEO re: current WGISS github account.	**Due: ASAP**
WGISS-57-04	WGISS Exec to convene a small group to tackle the topics around data replication in the cloud, using the CGMS Cloud Best practices document as a reference. Start with a GitHub repository and sketch an outline. Alex Leith and Matt Paget will join the group.	**Due: Q2 2024**
WGISS-57-05	Ken to ask his team about the relevance of the WGISS semantics in the NOAA Knowledge Graph work, or if there is NOAA semantics work that could be provided to WGISS to support the vocabulary (semantics) Interoperability Framework factor.	**Due: ASAP** **COMPLETE**
WGISS-57-06	Simon to brief Liping and Patrick via email about the shortcomings with the CEOS-ARD product specifications, as input for the ISO/OGC ARD Standards Working Group.	**Due: ASAP** **COMPLETE**
WGISS-57-07	WGISS Chair to compile a table of POC for external and internal collaborations. Table should be regularly	**Due: WGISS-58**

	updated (revised at least annually) and may be included as part of a WGISS annual work plan.	
WGISS-57-08	WGISS Chair to update the WGISS Terms of Reference (last updated 2015), specifically to address outdated Stakeholders section but other sections should also be reviewed as part of the update. An updated Terms of Reference will be presented at CEOS Plenary 2024.	**Due: Q3 2024**
WGISS-57-09	WGISS Chair to work with WGCV Chair to map out an outline for a Satellite Earth Observation Commercial Engagement Workshop at WGISS-58 / WGCV-54 in October 2024 modelled after the LSI-VC workshop to be held in April 2024 (using lessons learned from that workshop).	**Due: WGISS-58**
WGISS-57-10	Libby to work with Exec on developing a standard WGISS overview presentation. Each IG Chair should provide at least one slide describing activities within the IG. The Chair Team will provide slide(s) describing WGISS overall goals and recent accomplishments.	**Due: WGISS-58**
WGISS-57-11	Libby to update the brochure link online to the more recent version provided by Iolanda. Consider whether to remove the brochure once the slide deck in WGISS-57-10 is complete.	**Due: ASAP**
WGISS-57-12	DIIG to contact Brian Terry & the SEO team regarding the use of the CEOS Analytics Lab for the EO Plug and Play (PnP) demonstrators.	**Due: WGISS-58**
WGISS-57-13	Interoperability IG Lead to work with Katrin to incorporate components of Common Online Dictionary Research Paper into the Interoperability Handbook 2.0 Vocabulary Section.	**Due: WGISS-58**
WGISS-57-14	Libby to follow up with SEO regarding metrics for COVE use and access to WGISS Best Practices and Guides.	**Due: ASAP**
WGISS-57-15	DAIG to discuss with SEO the possibility of a central CEOS Authentication mechanism.	**Due: WGISS-58**
WGISS-57-16	ARD Oversight Group to identify where specifically WGISS should support the activities identified in the	**Due: Q2 2024**

	2024 CEOS-ARD strategy.	
WGISS-57-17	Interoperability IG Lead to set up a meeting with LSI-VC and ARD-Oversight Group leads regarding someone from those groups taking a leadership role for development of the Architecture component of the Interoperability Handbook 2.0.	**Due: Q2 2024**
WGISS-57-18	SEO (Alex Leith) to share information with WGISS about the SatSummit workshop coming up in mid-May 2024.	**Due: ASAP**
WGISS-57-19	WGISS Exec to discuss an approach to sharing data access success stories (such as Alex's code snippet demonstrating cloud native data access) and how we could facilitate instructions to stakeholders.	**Due: Q3 2024**
WGISS-57-20	Libby to work with TEIG on some communication materials around the Jupyter Notebook best practices.	**Due: Q2 2024**
WGISS-57-21	TEIG to contact EUMETSAT for some example notebooks to demonstrate the Jupyter Notebook best practices. Look at WGISS-56 presentations and FedEO catalogue.	**Due: ASAP** **COMPLETE**
WGISS-57-22	TEIG to work with SEO to ensure CEOS Analytics Lab notebooks are compliant with the Jupyter Notebooks best practices.	**Due: Q2 2024**
WGISS-57-23	TEIG to discuss a communications strategy for AI/ML white paper, to gather feedback and comments from the community. Present this to WGISS-58 for discussion.	**Due: WGISS-58**
WGISS-57-24	TEIG to develop [slide 5 from the ISRO presentation](#) into a figure available on the WGISS website.	**Due: WGISS-58**
WGISS-57-25	ISRO to coordinate with USGS & ESA on Landsat 7 image restoration work using AI/ML to infill missing data stripes.	**Due: WGISS-58**
WGISS-57-26	Liping Di to write up some feedback for CEOS Agencies about the use of EO data in CropSmart, in particular the search, discovery, access and exploitation aspects.	**Due: Q2 2024**

WGISS-57-27	Doug Newman to put Alex Leith in touch with the MAAP team for more details on the method they used to scale from a notebook to a global scale workflow.	**Due: ASAP**
WGISS-57-28	DAIG to share the invite for April and May STAC Best Practice calls with WGISS members, LSI-VC leads, and ARD Oversight Group lead.	**Due: ASAP**
WGISS-57-29	DAIG to consider updating the title of the STAC Best Practices to "Collection and Granule Discovery Best Practices with STAC", and proceed with Best Practices finalisation according to the plan presented by Yves Coene.	**Due: WGISS-58**
WGISS-57-30	DAIG to work with the CEOS Comms Team to put together an article for ceos.org/news regarding the new FedEO release and the retirement of CWIC.	**Due: Q2 2024**
WGISS-57-31	WGISS chair to work with SEO regarding the retirement of the website subdomain for CEOS IDN, and forwarding to the NASA IDN page.	**Due: Q2 2024**
WGISS-57-32	IDN team to review their POC list with respect to the list of WGISS principals, and confirm their representation in IDN. In particular, confirm the USGS contact. Set up a process to regularly (annually) ensure the POC is up to date, and any agency datasets represented in IDN are up to date.	**Due: Q2 2024**
WGISS-57-33	DAIG to add NASA on the connected data assets graphic, and strengthen the connection between IDN and FedEO. Potentially also add other WGISS "members", even if they are not currently uploading metadata to FedEO or IDN.	**Due: ASAP**
WGISS-57-34	DSIG to confirm with SEO to take down the Open Source Software Inventory, and make sure it is removed from the CEOS Website.	**Due: ASAP**
WGISS-57-35	TEIG to incorporate the recommendations from Yves in the Jupyter Notebook best practices.	**Due: ASAP**
WGISS-57-36	WGISS Chair to work with CEOS Executive Officer regarding document governance including definitions	**Due: WGISS-58**

	for best practice vs. white paper vs. case study, and at what level (WG vs. SEC/Plenary) each type of document should be approved.	
WGISS-57-37	WGISS Exec to review the New Agency Dataset page, and whether it is still used. The process for updates to this page should be clarified.	**Due: WGISS-58**
WGISS-57-38	DSIG to schedule a discussion for WGISS-58 about the various agency archiving policies. Consider developing a best practices document for archiving policies (and technologies).	**Due: WGISS-58**
WGISS-57-39	Libby to work with DSIG to draft up the collaborations page on the WGISS website.	**Due: ASAP**
WGISS-57-40	WGISS Chair to set up a meeting with Mirko Albani and USGS archive colleagues alongside WGISS-58 regarding the USGS AVHRR holdings (stitched data, documents, tools). Verify the USGS AVHRR holdings ahead of October.	**Due: WGISS-58**
WGISS-57-41	ISRO to check the availability of AVHRR raw data, and prepare plots showing time period availability of AVHRR data from each mission.	**Due: ASAP**
WGISS-57-42	DSIG to consider an annex to the DMSMM white paper containing a guide on how to use the maturity matrix.	**Due: WGISS-58**
WGISS-57-43	DSIG to invite NOAA to present on their maturity matrix COMET tool at WGISS-58, and consider whether WGISS should develop a similar tool.	**Due: WGISS-58**
WGISS-57-44	WGISS Exec to discuss how WGISS might be involved in the CEOS biodiversity activities, responding to the challenges presented at WGISS-57.	**Due: Q2 2024**
WGISS-57-45	GA and USGS to share their work on collection management to the broader group for review via GitHub (coordinate with Libby). DSIG will organise a meeting in May 2024 to discuss the work done by GA and USGS, and the way forward on a WGISS white paper.	**Due: ASAP**