MINUTES

OF THE

54th MEETING

OF THE

CEOS WORKING GROUP ON   
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Tokyo, Japan

October 4-6, 2022

Hosted by

Japan Aerospace Exploration Agency, JAXA

Table of Contents

[1 WGISS Plenary Session, Part I 7](#_Toc118731890)

[1.1 Opening Remarks and Logistics 7](#_Toc118731891)

[1.2 Welcome from the Director of JAXA’s Earth Observation Research Center 7](#_Toc118731892)

[1.3 WGISS Chair Report 7](#_Toc118731893)

[1.4 CEOS Executive Officer (CEO) Report 8](#_Toc118731894)

[1.5 CEOS Systems Engineering Office (SEO) Report 9](#_Toc118731895)

[1.6 GEO Secretariat Report 10](#_Toc118731896)

[1.7 ISO TC211 Liaison Report 11](#_Toc118731897)

[2 Data PRESERVATION and STEWARDSHIP 12](#_Toc118731898)

[2.1 Session Objectives 12](#_Toc118731900)

[2.1.1 WGISS DSIG Best Practices Refreshment 12](#_Toc118731901)

[2.1.2 FAIR Dataset Quality Information Guidelines 13](#_Toc118731902)

[2.1.3 AVHRR Recovery Activity 13](#_Toc118731903)

[2.1.4 Cooperation with CGMS 14](#_Toc118731904)

[2.1.5 Archive Technology and Future Trends 14](#_Toc118731905)

[3 Data DISCOVERY and ACCESS 19](#_Toc118731906)

[3.1 International Directory Network (IDN) Report 19](#_Toc118731908)

[3.2 Federated Earth Observations (FedEO) Update 19](#_Toc118731909)

[3.3 OGC Testbed-18: Secure, Asynchronous Catalogues 19](#_Toc118731910)

[3.4 Service Metadata and Discovery Best Practices 20](#_Toc118731911)

[4 Data INTEROPERABILITY and USE 22](#_Toc118731912)

[4.1 EO Interoperability for Services 22](#_Toc118731914)

[4.2 Session on EAIL: Use-cases and Vision 22](#_Toc118731915)

[4.2.1 CEOS Coastal Observations, Applications, Services, and Tools (CEOS COAST) 22](#_Toc118731916)

[4.2.2 CSIRO 23](#_Toc118731917)

[4.3 STAC 24](#_Toc118731918)

[4.3.1 Federated Discovery using STAC 24](#_Toc118731919)

[4.3.2 Access Patterns and STAC Implementation at Bhoonidhi 24](#_Toc118731920)

[4.3.3 STAC Best Practice 24](#_Toc118731921)

[4.4 Discussion 25](#_Toc118731922)

[5 TECHNOLOGY EXPLORATION 26](#_Toc118731923)

[5.1 Jupyter Notebooks 26](#_Toc118731925)

[5.1.1 Update on Jupyter Notebooks Day 26](#_Toc118731926)

[5.1.2 Jupyter Notebooks Best Practice Document 26](#_Toc118731927)

[5.1.3 WGCapD Updates and Webinar Support 27](#_Toc118731928)

[5.2 Machine Learning, Data Learning, Artificial Intelligence: Session Objectives 27](#_Toc118731929)

[5.2.1 NOAA Center for Artificial Intelligence 27](#_Toc118731930)

[5.2.2 Enhancing Earth Observation Applications Using AI/ML Techniques 28](#_Toc118731931)

[5.2.3 Machine Learning and Graph 28](#_Toc118731932)

[5.2.4 USGS AI/ML Use Cases 29](#_Toc118731933)

[5.2.5 New AI/ML Keywords Introduced to GCMD 29](#_Toc118731934)

[5.2.6 Introducing White Paper for AI/ML Outline 30](#_Toc118731935)

[5.3 Federation 30](#_Toc118731936)

[6 WGISS-WGCV Joint Symposium 32](#_Toc118731937)

[6.1 WGCV Overview 32](#_Toc118731939)

[6.2 WGISS Overview 32](#_Toc118731940)

[6.3 CEOS CAL/VAL Portal, ESA EO Portal, CEOS MIM Database: Status, Synergies, and Improvements 32](#_Toc118731941)

[6.3.1 ESA eoPortal and CEOS MIM Database 32](#_Toc118731942)

[6.3.2 CEOS CAL/VAL Portal 33](#_Toc118731943)

[6.3.3 Discussion 33](#_Toc118731944)

[6.4 Status of ARD Activities 33](#_Toc118731945)

[6.5 Data Quality Assessment and Indicators, and DMSMM Maturity Matrix 34](#_Toc118731946)

[6.5.1 Data Quality Indicators 34](#_Toc118731947)

[6.5.2 Data Management and Stewardship Maturity Matrix: Introduction, Background and Definitions 34](#_Toc118731948)

[6.5.3 AVHRR DMSMM Use Case 35](#_Toc118731949)

[6.5.4 Discussion 35](#_Toc118731950)

[6.6 CEOS Interoperability Framework Initiative 36](#_Toc118731951)

[6.7 CEOS Common Online Dictionary: Progress with Joint Activity between WGISS and WGCV 37](#_Toc118731952)

[6.8 Greetings from the Senior Chief Officer of Earth Observation Missions 38](#_Toc118731953)

[7 Agency and Liaison Reports 39](#_Toc118731954)

[7.1 Indian Space Research Organization (ISRO) 39](#_Toc118731956)

[7.2 Japan Aerospace Exploration Agency (JAXA) 39](#_Toc118731957)

[7.3 United States Geological Survey (USGS) 40](#_Toc118731958)

[7.4 German Aerospace Center (DLR) 40](#_Toc118731959)

[7.5 National Aeronautics and Space Administration (NASA) 41](#_Toc118731960)

[7.6 Agencia Espacial Mexicana 42](#_Toc118731961)

[8 WGISS Plenary, Part II 43](#_Toc118731962)

[8.1 Future Meetings 43](#_Toc118731964)

[8.2 WGISS Vice Chair Solicitation 43](#_Toc118731965)

[8.3 WGISS Summary and Discussion 44](#_Toc118731966)

[8.4 Review of WGISS Actions 45](#_Toc118731967)

[8.5 Concluding Discussion and Remarks 46](#_Toc118731968)

[9 Glossary of Acronyms 47](#_Toc118731969)

Agencia Espacial Mexicana Adrian Guzman\*

CAS, AIR Ziyang Li\*, Xuesong Li\*

CEOS Marie-Claire Greening\*

CNES Richard Moreno, Karim Ramage (CNRS)\*

CONAE Homero Lozza\*

CSIRO Jonathan Hodge\*

DLR Katrin Molch\*

ESA Mirko Albani, Roberto Alacevich\*, Pavan Alikana (Rhea)\*, Paolo Castracane (Rhea)\*, Yves Coene (Spacebel)\*, Anatole Deligant\*, Ferran Gascon\*, Philippe Goryl\*, Damiano Guerrucci\*, Daniele Iozzino (Rhea)\*, Brian Keary\*, Iolanda Maggio (Rhea), Filippo Marchesi (Solenix)\*, Paulo Sacramento\*

EUMETSAT Uwe Voges\*

European Commission Peter Strobl\*

GEO Paola de Salvo\*, Laurent Durieux\*

GISTDA Pakorn Apaphant, Sittiporn Channumsin\*, Napawan Panya\*, Panu Srestasathiern, Nuttavipa Thanthawewut

ISRO Nitant Dube\*, Ankitha E\*, Radha Kavuluru\*, Ankitha Reddy\*, Sai Kalpana Tanguturu\*

JAXA Makoto Natsuisaka (WGISS Chair), Riko Oki, Yukio Haruyama, Yousuke Ikehata, Kaori Kuroiwa, Akihiko Ku, Shinichi Sobue, Michihiro Koide, Michelle Piepgrass, (WGISS Secretary)

LSI-VC Matt Steventon (Symbios)\*, George Dyke (Symbios)\*

NASA Andrew (Andy) Mitchell, Lauren Childs-Gleason\*, Diane Davies, Liping Di (GMU)\*, Valerie Dixon, Brian Killough\*, Dawn Lowe, Michael Morahan, Doug Newman, Kenton Ross\*, Min Wong

NOAA Kenneth (Ken) Casey\*, Prasanjit Dash\*, Paul DiGiacomo\*, Heng Gu\*, Veronica Lance\*, Merrie Beth Neely\*, Redmon Robert\*, Nancy Ritchey\*, Michael Soracco\*, Julie Warner\*, Ivana Ivanova (Curtin University)\*

UKSA Esther Conway, Ed Williamson (STFC)\*

UNOOSA Jorge del Rio Vera\*

USGS Tom Sohre (WGISS Vice-chair)

**GROUP PHOTO**

****

# WGISS Plenary Session, Part I

## Opening Remarks and Logistics

Makoto Natsuisaka, WGISS-Chair (JAXA) opened the WGISS-54 meeting, thanking everyone for their participation, He asked the participant to introduce themselves, and reviewed the agenda. He noted that this is a working meeting, and asked in-person and remote attendees to participate.

## [Welcome from the Director of JAXA’s Earth Observation Research Center](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_09.45_JAXA_EO_Program.pptx)

Dr. Riko Oki (JAXA), director of Director of JAXA’s Earth Observation Research Center, welcomed the participants to Japan and the first in-person meeting after the Covid-19 pandemic. She described the JAXA EO Program organizational chart, noting that Hirabayashi Takeshi is the Senior Chief Officer of Earth Observation Missions, and the CEOS JAXA principal. Two primary divisions are EORC for research, calibration, and validation and SAOC for satellite operation data production and distribution.

The Japan Aerospace Exploration Agency (JAXA) was born through the merger of three institutions, namely the Institute of Space and Astronautical Science (ISAS), the National Aerospace Laboratory of Japan (NAL) and the National Space Development Agency of Japan (NASDA). It was designated as a core performance agency to support the Japanese government's overall aerospace development and utilization. JAXA, therefore, can conduct integrated operations from basic research and development to utilization.

In 2013, to commemorate the 10th anniversary of its founding, JAXA created the corporate slogan, "Explore to Realize," which reflects its management philosophy of utilizing space and the sky to achieve a safe and affluent society.

Dr. Oki listed the past, current, and planned missions, and projects GCOM, EarthCARE/CPR, GOSAT-2, ALOS-3, ALOS-4, and GOSAT-GW. She noted that satellite temporal coverage is the biggest challenge of satellite observations, and to compensate JAXA seeks international cooperation by participating in seven CEOS entities. Owing to the WGISS activities, JAXA benefits from the WGISS Connected Data Assets, as well as the other activities of WGISS.

Dr. Oki expressed the hope that participants will take advantage of the in-person meeting, have fruitful discussions, and enjoy Tokyo.

## [WGISS Chair Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_10.00_WGISSChairReport%20.pptx)

Makoto Natsuisaka (JAXA) gave a report summarizing the work of WGISS since the WGISS-53 meeting. Major events include:

SIT37 (March 29- 31, 2022)

SIT-TW (September13-15)

CEOS SEC meetings

CEOS deliverable update toward SIT-TW and Plenary

CEOS ARD Oversight Group

CEOS Ocean Coordination Group

CEOS Interoperability Framework

CEOS Engagement with Standard Organizations

LSI-VC-12 (September 9)

CEOS activities in 2022

Chair CNES initiative “Paths to Sustainability: from strategy to practical measures Version 1.0 – 25 October 2021” in response to the Paris Climate Agreement, the Sendai Framework for Disaster Risk Reduction, and the 2030 Agenda for Sustainable Development.

The engagement with the International Methane Emissions Observatory (IMEO) launched at the G20 Summit is a data-driven, action-focused initiative by the UN Environment Programme (UNEP) with support from the European Commission is one of the new topics. <https://www.unep.org/explore-topics/energy/what-we-do/imeo>

Cooperation with “New Space” is also one of the new topics.

Makoto reported that the CEOS ARD Oversight Group is identified as a core component of the future CEOS ARD Governance Framework, to act as a forum for all matters related to CEOS ARD. Representatives from the CEOS Virtual Constellations are integral to the effort going forward, as these are the CEOS entities with the technical expertise to recommend, develop and maintain Product Family Specifications. The group also provides strong coordination across CEOS and promotes CEOS ARD in a unified way

The new CEOS Ocean Coordination Group is a task force led by SIT Chair to coordinate CEOS ocean-related activities and clarify CEOS contribution to UN Ocean Decade. COAST expects the EAIL continuity and has an interest AI/ML.

Makoto cited the cooperating relationships and meetings with other CEOS entities:

CEOS ARD Oversight Group (CEOS SIT)

CEOS Ocean Coordination Group (CEOS SIT)

CEOS Interoperability Framework (LSI-VC) 🡪 WGISS-54 joint session

CEOS Engagement with Standard Organizations (LSI-VC) 🡪 WGISS-54 joint session

EAIL initiative … SEO, WGDisasters (Flood Pilot Project), WGCV (DEMIX), COAST, Rice Monitoring Community 🡪 WGISS-54 Interoperability and Use session

Jupyter Notebook initiative (SEO, WGCapD) 🡪 WGISS-54 Technology Exploration session

Common DSMM (WGCV, CGMS group IV) 🡪 WGISS-54 Data Stewardship session

Common On-line Dictionary (WGCV) 🡪 WGISS-54 joint session

ISO TC211/OGC🡪 WGISS-54 liaison report

Makoto listed agenda highlights of WGISS-54. He noted that CEOS Interoperability Framework and CEOS Engagement with Standard Organizations would be discussed during the joint session with WGCV. He added that a task force which consists of LSI-VC, WGCV, and SEO is preparing a ToR and will submit it to the CEOS Plenary to be held in November.

Makoto concluded saying that he expects a review of the WGISS Organizational Structure and the WGISS ToR, making potential updates. Specifically, the Data Interoperability and Use Interest Group may need to be reorganized, in preparation for WGISS contributions to the CEOS Interoperability Framework and the CEOS Engagement with Standards Organizations initiatives.

WGISS is also formally requesting nominations for WGISS Vice-chair (beginning November 2023), and confirmation of future meetings.

## [CEOS Executive Officer (CEO) Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_13.30_CEOReport.pptx)

Marie-Claire Greening (CEO) gave a report on CEOS activities. She mentioned the CEOS mission, objectives, and long-term priorities. She also reiterated the 2022 Chair priorities, highlighting the close collaboration of CEOS with other global entities. She noted that the WGISS mission and objectives enable the objectives of CEOS.

Marie-Claire reminded of the four key governance documents, of which the CEO is custodian:

1. [CEOS Terms of Reference](http://ceos.org/document_management/Publications/Governing_Docs/CEOS_Terms-of-Reference_Nov2013.pdf) defines the mission and scope of CEOS activities.
2. [CEOS Strategic Guidance document](http://ceos.org/document_management/Publications/Governing_Docs/CEOS_Strategic-Guidance_Nov2013.pdf) articulates the overarching long-term (7-10 years) purpose and goals of CEOS.
3. [CEOS Governance and Processes](http://ceos.org/document_management/Publications/Governing_Docs/CEOS_Governance_and_Processes_rev1.1-2019.pdf) document provides guidelines on the structure, operations, and processes CEOS employs to achieve its goals.
4. [CEOS Work Plan](https://ceos.org/document_management/Publications/CEOS_Work-Plans/CEOS_2021-2023-Work-Plan_Mar2021.pdf) (3-year rolling) sets forth near-term actions to achieve the goals outlined in the CEOS Strategic Guidance document. Detailed work is defined as deliverables, outlined in the WP, reconciled, and tracked in [CEOS deliverable tracking tool](http://deliverables.ceos.org/).

The CEOS Work Plan is structured with 12 chapters. WGISS is responsible for chapter 9: Data discovery, access, preservation, useability, and exploitation: approaches, systems, tools, and technologies. This chapter contains three deliverables due at the end of 2022, and three due at the end of 2023. WGISS needs to define deliverables for 2024, and it is preferred that they be submitted in the first two months of 2023.

**Action WGISS-54-01: Makoto Natsuisaka and WGISS Exec to begin drafting a list of CEOS Workplan deliverables for the period 2023-2025. Due January/February 2023.**

## [CEOS Systems Engineering Office (SEO) Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_10.20_SEO_Report.pptx)

Brian Killough (NASA) gave a report on SEO activities. He reported that the SEO has been testing several cloud computing frameworks to understand CEOS data access and technology capabilities. Though these services seem identical, how they operate in the background is quite different and much has been learned about the advantages and disadvantages of each.

Google – using the Google Cloud (paid compute environment), Colab (free notebook platform, but limited) and Earth Engine (free satellite datasets).

Amazon – using the AWS Cloud (paid compute environment), SageMaker (free notebook platform, but limited), and the AWS Open Data Catalog (free satellite datasets).

Sentinel Hub –working with Sinergise to test ODC integration with the Sentinel Hub via CreoDIAS (European cloud provider).

Microsoft –started testing with the Microsoft Azure Cloud (paid compute environment), Azure Labs (free notebook platform, but limited) and the Planetary Computer Data Catalog (free satellite datasets).

The Open Earth Alliance (OEA) is a new GEO Community Activity created in 2021 by the ODC Founding Partners to expand the impact of ODC, support the concept of regional data cubes, and explore new open-source data solutions and technologies. The SEO and CSIRO have been the primary contributors to this initiative. Microsoft Azure GEO award credits have supported testing and integration of ODC using Jupyter Hub (notebook server) and Dask (parallel processing). Kubernetes containers are also being tested for efficient deployments. The initial results are very promising as these solutions are 100% open-source and provide scalable data science solutions. Brian noted that data science and GIS in general are merging quite rapidly.

The SEO works with three regional data cubes:

Digital Earth Africa – The SEO led the deployment and testing of the African Regional Data Cube in 2018. The success of this prototype led to a fully funded DE-Africa. Leadership has now completely transitioned to Africa (SANSA). The SEO still represents CEOS on the Steering Group.

Digital Earth Pacific – The SEO has worked with the Pacific Community (SPC, formerly the South Pacific Commission) to guide the initial prototype and conduct user needs meetings. SPC is now working on securing funding. The SEO represents CEOS on the Steering Group.

Digital Earth Americas – The SEO is working with CSIRO (Chile) and INEGI (Mexico) to explore opportunities toward a future DE-Americas. To date, several workshops have been conducted in the region and explored user needs. These workshops have involved several regional stakeholders (e.g., AmeriGEO, ECLAC). A future Caribbean initiative funded by Mexico is likely the next opportunity to build awareness and explore.

The Open Data Cube (ODC) Sandbox now has 19 applications in GitHub. This open-source tool is a great example of “open science” and has gained significant popularity as an educational and research tool. The SEO worked with Digital Earth Africa to develop many of their ODC applications and tools. They recently released 100+ application notebooks online focused on satellite datasets, real-world examples, and SDGs.

The SEO is planning to participate in several webinars to promote open science and use of CEOS satellite data. These include WGCapD “Jupyter Notebook Day” in late October and the GEO Knowledge Hub (future 3-part webinar series). Each event will use the ODC Sandbox.

The Earth Analytics Interoperability Lab (EAIL) continues to operate and support CEOS projects. Seventeen data product pipelines are available (Landsat 5-9) Surface Reflectance and Temps, Sentinel-1-RTC, Sentinel-2, and DEM (NASA, Copernicus). Jonathan Hodge (CSIRO-Chile) is the primary lead and architect of EAIL with some support from the SEO. EAIL operates on the Amazon Cloud using Open Data Cube and supports Dask scaling for larger analyses. The EAIL is currently supporting data and algorithm testing for several CEOS initiatives e.g., COAST (in progress) and WGCV (in early planning to support Cal-Val campaigns).

**Action WGISS-54-02: WGISS Exec to request from the SEO a copy of the charter that describes the goals and objectives of cloud prototyping efforts. If a charter does not exist, create a follow-up action to work with SEO to charter the efforts so that CEOS has defined goals and objectives.**

Discussion points:

It would be helpful to understand the goals and objectives of SEO cloud prototyping efforts. Documenting best practices and lessons learned should be an object. NASA has a large body of these that could be consolidated and made available. It would be helpful to provide guidance in how to set up an environment and make it interoperable among various clouds. Federation and discoverability should also be considered.

CEOS members can send service and tool records to enter in the CMR. As preferred/interoperable services are found and shared the IDN team could add UMM-Service/Tool capability to the draftMMT tool so CEOS members can add and update service and tool metadata records themselves. Services and Tools can be associated to data collections that they operate upon or are compatible with, making it easier for the IDN end user to tailor or perform analysis upon selected data.

## [GEO Secretariat Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_17.15_GEO%20Secretariat%20Report.pdf)

Paola de Salvo (GEOSEC) gave a report on GEO activities related to data and knowledge. She reported that the GEO Data Working Group (DWG) is focusing on three main aspects: in-situ data, data sharing (data management principles), and law and policy. They conducted a GEO Work Programme activities survey to assess the GWP activities, data-related practices, challenges, actions, assess their knowledge about the GEO Data Sharing and Data Management Principles, and needs related to the GEOSS Platform. Launched in October 2021, 77% of respondents stated that they are implementing the DMPs fully or to a certain extent.

Following the survey, the DWG conducted engagement calls to discuss:

* Inventory of data needs (unavailable data, restrictions, formats, accessibility) and key networks with which to engage.
* Understanding of awareness about the Data Sharing and Data Management Principles and the support needed for their implementation.
* Understanding of the law and policy about data used by the activities.
* Gathering feedback about the infrastructure needs for the enhancement of the GEOSS Platform.

The GEO Data Sharing and Data Management Principles subgroup is running a webinar every Monday from 16:00 to 17:30 CEST to raise awareness about the GEO principles and their benefits. The webinars are aimed at all Earth observation stakeholders, including data producers, technology providers, scientists, researchers, business developers, decision makers and policy makers. Members of the Earth observation community will discuss their experience implementing the principles, talk through how they tackled challenges and showcase the impact they realized.

Paola presented the planned roadmap of the GEO Knowledge Hub and gave a live demonstration of the hub, an operational Platform with User and Knowledge Provider customized Interface and Functionalities, 30 Open EO Applications discoverable and re-usable, link with 2 EO ToolKits (Disasters and Urban), and contribution from 12 GEO Work Programme activities with many more in the pipeline.

Current planned goals are to

Advocate GKH uptake through an integrated Users strategy

Continue the engagement with GEO WP Activities (30 new knowledge packages in the pipeline)

GKH as impact tool integrated in the National GEOs engagement strategy

Dedicated Open data and Open Knowledge workshop in 2023

The Data Working Group, the GEOSS Platform, the GEO Knowledge Hub are advocating the need to have a technical event to gather the GEO community in 2023. Discussions are ongoing to host it in ESA in March 2023. Initial ideas were listed.

Regarding the GEO Portal connection to the IDN Paola said that there is no direct connection, but for the GEOSS Platform there is. GEO are seeking integration depending on the output of the Expert Advisory Group, who are working on strengthening the integration of GEOSS Platform and GEO Knowledge Hub. The goal is not to replicate the data, but rather to have the systems work in a synchronized manner. A connection of IDN with GEO Knowledge Hub is planned; there are also plans to have a series of training for the community who would benefit from WGISS knowledge and experience from these technological advancements.

Iolanda noted that WGISS-54 discussions include some ideas for improvement of the self-assessment tool for the Data Management and Stewardship Maturity Matrix which is based on the GEO Data Management Principles. She wondered what the status of the tool in the GEO Knowledge Hub is. Paola replied that the self-assessment tool is in the GEO Knowledge Hub, and will be discussed 24th of October; it would be great if WGISS could join the conversation that day, to broaden the discussion and to see if the tool is comprehensive enough. Iolanda said that she could join, and would like to share experience. Paola will follow-up with Iolanda and the authors from the GEO side to coordinate.

Makoto mentioned the upcoming Jupyter Notebook Day October 21 and Esther agreed to share the information with Paola to distribute as appropriate.

**Action WGISS-54-03: Iolanda Maggio and Mirko Albani to participate in the discussion with GEO on 24th of October regarding the self-assessment tool in the GEO Knowledge Hub, to share experiences based on the WGISS Data Management and Stewardship Maturity Matrix which is based on the GEO Data Management Principles.**

## [ISO TC211 Liaison Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_10.40_ISO%20TC211%20Liaison%20Report.pptx)

Liping Di (GMU for NASA) gave the ISO TC 211 Liaison report, giving the scope and status of ISO TC 211 standard projects related to CEOS/WGISS:

* ISO 19123-1 Geographic information — Schema for coverage geometry and functions — Part 1: Fundamentals. Passed Draft International Standard vote. Editing Committee Meeting was held in August 2022. Currently in Formal Draft International Standard (FDIS) Stage.
* ISO 19124-1 Geographic information – Calibration and Validation of Remote Sensing Data and Derived Products – Part 1: Fundamentals. Passed the Draft Technical Specification (DTS) vote. Editing Committee meeting will be held in Oct 2022. Will be advanced to TS after editing.
* ISO 19159-4 ISO 19159-4 Geographic information — Calibration and validation of remote sensing imagery sensors — Part 4: Space-borne passive microwave radiometers. At TS stage. To be published soon.

At OGC Disaster Pilot 2022, OGC, by working with CEOS, developed a plan to make ARD a joint ISO-OGC standard:

* Set a standard working group (ARD SWG) in OGC to develop ARD standards
* Start an ISO TC 211 project on ARD by submitting a new work item proposal through OGC-ISO JAG
* The ISO TC 211 project team and OGC ARD SWG will have the same group of subject-matter experts
* ARD series of standards will be a multi-part standard, and both OGC and ISO version will be the same

The New Work Item Proposal (NWIP) will be discussed at ISO TC 211 55th Plenary meeting in Stockholm, Sweden. The NWIP will be voted by TC211 member countries in the following months. However, lack of resources to support the work is the key issue.

The ISO ISO/TC 211 55th Plenary meeting is December 5th - December 9th. Does CEOS WGISS want to report anything to ISO TC 211? Inputs must be provided to Liping by November 20.

Ken Casey asked why ARD will be a multi-part standard? Liping replied that this follows the ARD structure, governance, and framework. For different types of sensors (optical, lidar, microwave, SAR) there are different methods of calibration, different mandatory processes. If put all together, the document will be too large and be confusing.

Makoto highlighted that his work is voluntary at this time.

# Data PRESERVATION and STEWARDSHIP



## [Session Objectives](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_11.20_DSIG%20Session%20Introduction.pptx)

Mirko Albani (ESA) discussed the objectives for the Data Preservation and Stewardship session. He noted that WGISS accomplishes its Data Preservation and Curation efforts through the Data Stewardship Interest Group (DSIG) with a Focus on EO Data and Technical Content.

Mirko listed the technical documents produced to this date. These include general and specific guidelines and best practices, and technical implementation procedures (such as the Purge Alert Service White Paper). The EO Data Preservation Guidelines support the “Detailed Definition” and “Implementation” steps of the Stewardship Reference Model.

The DSIG organizes technical sessions on various topics like Archive Holdings and Technology, Data Provenance, and Heritage Data Recovery.

### [WGISS DSIG Best Practices Refreshment](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_11.45_DSIG%20Best%20Practices%20Refreshment.pptx)

Iolanda Maggio (RHEA for ESA) discussed a review of the WGISS Best Practices delivered by the DSIG. The goal is to introduce new terms occurring due to the rapid technology evolution, updated references, harmonization of statements and terminology, and updated use case, recommendation, and guidelines. Iolanda listed the DSIG BP landscape, which can be found on the WGISS Preservation and Stewardship web page.

1. EO Preserved Dataset Content: This is the starting point for ISO/DIS 19165-2 Geographic information — Preservation of digital data and metadata —- Part 2: Content specifications for earth observation data and derived digital products. Lessons Learned: using it in the ESA Preservation Workflow task the Data Curator identified few additional notes to be added to the BP. Should these improvements be communicated to Hampapuram Ramapriyan or can the team create an internal version 1.1 to WGISS? Since the ISO process takes 7 years, it was suggested that it be sent to Rama, but not wait for ISO approval.
2. EO Data Preservation Guideline: Only minor changes in the THEME 7: DATA EXPLOITATION AND RE-PROCESSING - GUIDELINE 7.12 – Facilitate data citation - (Level B): Landing page concept needs to be added.
3. Preview Image Principle: this concept is already in other BPs. Iolanda proposed to remove it from the website. After discussion, it was suggested that the whole concept of browse be readdressed.
4. Data Management Statement: Is it a real BP? Could be removed? The included concepts are already present in other BPs. If some original statements are found, they will be included in the relevant BP. Since the statement is within all BPs, it was agreed that it was not needed here.
5. CEOS Persistent Identifier Best Practices: Using this BP in the ESA context a few comments were collected, such as broken web links, agency-related technical aspects should be removed because each DOI registration agency implements a specific interface. On Landing Page following the statement of Best Practices for DOI Landing Pages from Datacite (<https://support.datacite.org/docs/landing-pages>). It was agreed that there is room for improvement on the landing page guidelines.
6. Generic Earth Observation Data Set Consolidation Process: no changes identified
7. Long term Preservation of EO: Preservation Workflow contains all detailed tasks of the Reference Model White Paper. The document needs a refreshment in terms of terminology and a verification of the links inside. It was agreed that the minor improvements should be implemented.
8. Long term Preservation of EO: Glossary of Acronyms and Terms. This is one of the starting points for the CEOS Online Dictionary. Should this be aligned with respect to the CEOS Online Dictionary? Or it will be superseded by the CEOS Online Dictionary? This glossary has already been submitted to the CEOS Online Dictionary team.
9. CEOS Technical Content and Associated Information Preservation Best Practices: A few formats for long term preservation of associated information should be updated considering new trends and standards. The team proposes separation of the Associated information BP and SW preservation annex that can be considered as a technical white paper. The SW preservation white paper could be focused and updated with new technical solutions. The suggestion was accepted, noting that the BP should be done first.
10. Measuring Earth Observation Data Usage Best Practices: New metrics suggested; a couple of metrics not fully clear or might be redundant with others.
11. White papers can also be refreshed.

Having collected agreements concerning the implementation of the proposed refreshments, an implementation plan will be circulated during October 2022. Work will start immediately.

**Action WGISS-54-04: Data Stewardship Interest Group to issue a plan for data stewardship best practices refreshment with priorities and schedule. Due by October 2022.**

**Action WGISS-54-05: WGISS Exec to review prior browse data best practices and guidelines, and consider updating to reflect newer cloud capabilities.**

### [FAIR Dataset Quality Information Guidelines](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_12.15_FAIR_Dataset_Quality_Information_Guidelines.pdf)

Ivana Ivanova (Curtin University for NOAA) discussed the FAIR dataset information guidelines. She noted that the FAIR DQI community guidelines provide specific advice on ensuring quality metadata is compliant with the FAIR principles for the dataset. The FAIR DQI community guidelines are a living document developed by and for the international community. Use cases on challenges with quality information are wanted. FAIR DQI guidelines support Priority 3: Support to CEOS Cal/Val Initiatives to increase CEOS Agency Cal/Val Collaboration.

Ivana observed that the reuse of a dataset, particularly where multiple datasets are being merged, requires knowledge of the “quality” of the datasets especially where datasets are repurposed for use cases beyond what the original creator intended. With the rise of Artificial Intelligence (AI) and Machine Learning (ML), a new interpretation of FAIR is that it stands for “Fully AI Ready”: knowing the “quality” of data to be used is essential to avoid erroneous conclusions.

An international effort has come together to develop “community-agreed” guidelines on quality. Several journal articles have been published. Guidelines for development principles include taking a whole dataset-lifecycle approach; being quality-attribute and assessment-type agnostic; common terminology is essential for enabling interoperability; developing for the community by the community through an iterative process and leveraging the experiences and expertise of a team of interdisciplinary domain experts and community best practices and standards.

The framework is defined by four dimensions: services, science, product, and stewardship.

The FAIR DQI guidelines are:

Guideline 1: Describing Dataset (e.g., version, producer) - Ensure the dataset is findable and accessible

Guideline 2: Utilizing a quality assessment model - Ensure the assessment model is findable and accessible.

Guideline 3: Capturing the assessment method and results - Ensure the quality information is interoperable and reusable (machine end users).

Guideline 4: Describing the assessment method, workflow, and results - Ensure the quality information is findable, accessible, citable, and reusable (human end-users)

Guideline 5: Reporting the dataset quality information - Ensure the information is FAIR

Use cases are being collected.

Mirko noted that this presentation would have been useful in the joint session with WGCV. He will present the outcomes and have some discussion; the team is very interested in the future evolutions.

### [AVHRR Recovery Activity](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_12.00%20AVHRR%20International%20Cooperation%20Activities.pptx)

Mirko Albani (ESA) gave a presentation on AVHRR international cooperation activities and partners.

The AVHRR cooperation objectives are to unfold and make accessible 1km AVHRR data from regional archives (possibly open and free), transcribing unique AVHRR data from heritage media and identifying a common format for AVHRR Level-1b data and pursue (re)processing from AVHRR data owners/holders and facilitating data discovery through the WGISS Connected Data Assets Infrastructure.

Since WGISS-53, the inventory of existing national/regional HRPT and LAC data archives is under consolidation. Identification of high priority heritage media is to be transcribed. Pilot activity started by ESA with Memnon (Sony Europe) to assess feasibility of transcribing three types of Optical Disks containing unique AVHRR data: successful transcription of the provided samples; full transcription of 300 disks from UK university of Reading plus additional available at ESA will be attempted in 2023. Data discoverability and reprocessing to be addressed at a later stage.

**Action WGISS-54-06: Mirko Albani to organize a session at WGISS-55 on AVHRR data recovery including dedicated presentations on European and North America LAC Archives and progress on other worldwide archives. Due by WGISS-55.**

### [Cooperation with CGMS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_13.50_Cooperation_with_CGMS.pptx)

Mirko Albani (ESA) gave a report on WGISS cooperation with the Coordination Group for Meteorological Satellites (CGMS), which globally coordinates meteorological satellite systems. This includes protection of in-orbit assets, contingency planning, improvement of quality of data, support to users, facilitation of shared data access and development of the use of satellite products in key application areas.

CGMS Working Group IV addresses data access and end user support. He presented the CGMS organizational structure, and discussed the scope of the CGMS Working Group IV, which works on data access and is thus closely related to the work of WGISS. Mirko listed the CGMS publications and Best Practices documents, and noted that WG-IV adopted the CEOS WGISS DSIG BP/guidelines as applicable to CGMS members datasets about three years ago. WGISS keeps them informed about future evolutions.

WGISS DSIG gave a presentation at the CGMS WG-IV meeting (April 22), which was well received.

WG-IV members are interested in the applicability of the DSIG guidelines to ground-based data and data stored on the cloud; this is a potential area of cooperation and BP/guidelines extension. WG-IV also expressed interest in a joint meeting or session with WGISS to share experiences and present respective activities including topics such as data access, cloud services, metadata.

Mirko proposed that the WGISS chair to contact CGMS WG-IV chair to discuss/identify opportunity for joint session/meeting. The WGISS membership agreed that it would be valuable to have joint meeting with CGMS and to foster this potential area of cooperation.

Nitant Dubhe commented that guidelines for ground-based observations would be a good discussion point with WGCV.

**Action WGISS-54-07: Mirko Albani to contact CGMS rapporteur (Simon Elliott, EUMETSAT) to understand if/how a join session with WGISS could be organized. (Note: CGMS WG-IV will have dedicated meetings in January 2023 and March 2023; CGMS chair is Kotaro Bessho, kbessho@met.kishou.go.jp). Due by October 2022.**

**Action WGISS-54-08: Data Stewardship Interest Group to evaluate potential extension/application of existing WGISS data preservation guidelines to ground-based (in-situ) data. Due by WGISS-55.**

### Archive Technology and Future Trends

#### [NOAA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_11.30_NOAA%20Archive%20Technologies.pptx)

NOAA Nancy Ritchey (NOAA) and Kenneth Casey (NOAA) gave a presentation on NOAA’s archive technology.

NOAA/NESDIS is building the NESDIS Common Cloud Framework, which is a Cloud enterprise architecture that will bring the data, processing, and dissemination into the a common, secure, and scalable architecture.

The high-level concept includes secure ingest of data and going through consolidated ingest, into storage, metadata management, and archive that is leveraged for algorithms in a compute environment as well as science exploration. There is also distribution and access of data.

Underlying this new cloud archive is a new, essential component: a knowledge graph (KG). The KG starts with data acquisition, then extract features from that data/metadata and align them to concepts and relationships to be stored in the graph. The knowledge graph is an integrated component of the archive workflow, it is not bolted on after the fact. The information and relationships are extracted throughout this workflow, and all these pieces are part of the data model.

The team are focused on deploying the minimum viable product as an initial operating capability by March of 2023. From there development of the archive functionality will continue while starting to migrate data into the workflow, leveraging the migration to inform the development of the enterprise archive and access functionality. While the full migration is expected to be complete about 2027, the ability to access data from the Cloud archive and leverage the technology will be available incrementally as the data is migrated into the Cloud.

The presentation concluded with a summary of recent activities at NOAA’s National Centers for Environmental Information (NCEI).

Discussion points: Valerie Dixon asked if NOAA is interested in using the CMR? Ken replied that they will be, and will be replacing their legacy systems with the cloud. Doug asked how many entities are in graph? Ken replied that this is an incremental building process. Yes, they plan to migrate the PDA functionality to the cloud. Can the KG be used to discover data? No decision yet if the KG will be exposed to users; will be working incrementally on this, and performance challenges will need to be addressed. Doug noted that NASA has been using it to enable data discovery for related collections and is very interested in providing low level access; the expertise of the user community is essential for taking advantage of the knowledge discovery. Richard noted that KG can be a part of interoperability, a better way to describe the datasets for the users. Ken noted that the ability to link across from one KG to another is an exciting feature.

#### [USGS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_14.15_USGS%20Archive.pdf)

Tom Sohre (USGS) gave a presentation on archive technology at USGS. He began by listing USGS policies and applications and tools, assessments, and roadmaps. The USGS data management goals are to provide data management, access, archive, and distribution for all datasets within the USGS historical archives that have long term relevance to science and support the USGS mission. They also seek to improve access to the land archive, utilize consistent data management approaches across all datasets, and develop and maintain data access, preservation, and distribution infrastructure to support the mission and other projects. Tom described the USGS data management process, data archiving, and the tape library system.

USGS Cloud methodology consists of a “Hybrid” approach for processing and distribution. Collection level processing utilizes cloud infrastructure, and forward processing relies on EROS hardware. Collection processing takes weeks instead of months with forward processing sitting next to the archive. New acquisitions (<=90 days) are delivered on-premises, and legacy collections are distributed via the cloud copy.

Future archive trends are to continue to evolve Cloud capabilities though Cloud egress costs are large and archive replication is challenging. Several end-users (both academia and commercial) continue to replicate the USGS archive for utilization and redistribution. Challenges exist regarding community awareness on legitimacy and differences contained within these remote collections. Investigations are underway to provide the redistributors with authenticated or certified status, and looking at services to compare checksum hash values back against the USGS holdings for validation.

Archive is managed and preserved using three copies: nearline, offline, and offsite. Currently, only USGS Trusted Digital Repositories (TDR) are realized as authoritative data holders. Tom asked if, given today’s IT services, are there new ways to maintain long-term preservation while still meeting programmatic requirements?

Regarding managing interim data versions, and since advancements in sensor models and calibration/validation continue to drive product improvements: How should these versions between major processing campaigns be handled? What retention and long-term preservation requirements exist? How are other organizations handling this challenge?

Discussion points:

Mirko noted that at ESA they are having the same challenges with versions; they keep the current and the previous versions. This challenge (multiple versions, their authenticity and provenance) is a point WGISS may address in future, discussing the related goals and options for technology. It was suggested to begin with a defined use case covering why it should be addressed now, with a good definition of the problem, clear challenges, and leading to a practical outcome. One possible use case would address the need to ensure the processing software is authenticated, given the trend to open-source science, though that may already be solved with the open-source software methodology. Andy recommended a dedicated session at a future WGISS meeting where others not in WGISS could participate.

Richard noted that, in the case of Sentinel, there is a new architecture, where the long-term archive will only keep the level 0 data, producing level 1 and 2 on demand as the less costly alternative.

Tom commented that a possible use case is a Commercial Cloud Service takes the whole archive, ingests it into their process, but there is no way for the user to know that this is authentic Landsat data.

Richard commented that, in Copernicus, the authenticity and provenance system has a method to check if changes have occurred.

Esther commented that it would be useful to address this also in the Jupyter Notebooks area, and to also address third-party distributors.

Doug commented that on GitHub, the software is deemed authentic because that is where it is.

**Action WGISS-54-09: WGISS Exec to define a use case that clarifies the challenges of authenticating data replicas in the cloud, including related goals, options, and technologies. Due by WGISS-55.**

**Action WGISS-54-10: Mirko Albani to organize a dedicated presentation at WGISS-55 on Copernicus Sentinels Ground segment architecture and authenticity and provenance approaches, inviting outside participants. Due by WGISS-55**

#### [NASA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_14.30_EOSDIS%20Data%20Archival.pptx)

Valerie Dixon (NASA) gave a report of data archival at NASA EOSDIS. She described the distributed active archives and the conceptual cloud-based architecture. The Earthdata Cloud system is a “Managed” commercial cloud for EOSDIS on Amazon Web Services that improves the efficiency of NASA’s data systems operations. It was initially architected for applications and mission data ingest, archive, distribution and has been operational since July 2019.

Valerie displayed a graph of cloud holdings metrics this year.

The Earthdata Cloud Platform is a multi-account, Infrastructure-as-a-Service (IaaS) cloud platform operating on Amazon Web Services under a single ESDIS owned top level “payer account,” providing shared cloud services and controls to EOSDIS.

The EOSDIS Cumulus Project is a cloud-based framework for data ingest, archive, distribution, and management and includes Harmony, NASA’s Earthdata Cloud Services System, which allows users to produce analysis-ready data by subsetting, reprojecting, and converting data to a cloud-optimized format. The Shared Services reduce redundancy. Cumulus includes the standard stewardship features.

Cumulus API and dashboard provides archive monitoring and management, backup and recovery that is in-cloud and tape-based.

Discussion points:

Tom commented that this is very similar to what USGS is doing.

Andy added that they are starting to write user stories, anecdotal stories for improved user understanding. Doug gave the example of an earthquake in Alaska where scientists could write their research papers with data gathered and analyzed within days.

Andy also wondered, with moving everything to the cloud, what is a BP for metrics, and was can be quantified as measures of success. Tom added that USGS is also struggling with finding a good metric that is obtainable. This could be an area of interest for WGISS.

**Action WGISS-54-11: WGISS Exec to study an approach to define and obtain useful and valid metrics of agency data use and access in the Cloud that quantify measures of success. Due by WGISS-55.**

#### [CNES](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_14.45_CNES%20archive%20technology.pptx)

Richard Moreno (CNES) gave a presentation of archive technology at CNES. He reported that CNES is currently installing a new Datalake in its Toulouse premises, which will embark the next CNES long term archive. He continued with a focus on the technologies and principles used for Long-term Archive.

The datalake contains new spatial and scientific data storage to improve data analysis in CNES. For CNES internal datasets there is centralised storage solution for all projects with address volumetry ramp-up for new project like SWOT, GeoDataHub, etc. For external datasets there is centralised storage to cache (copy) useful datasets like the Sentinel products (PEPS platform), and auxiliary data (surface or topography models, weather, orbit, etc.). The datalake allows data sharing between projects without duplication, and provides online datasets for real-time usage, web access (not only diffusion), parallel processing, and DataLabs. There is more collaboration and data sharing between institutions and more interoperability through a standardised Object API (S3) with Glacier features.

The Datalake is designed for high volume storage. All users access centralised on the S3 Disk system. S3 API on Datalake / POSIX only on the Application or processing. Data needs to be retrieved on the datalake and process on HPC or VM POSIX storage if the user needs POSIX semantic. The long-term archive solution has a specific API developed as a gateway between ground segment system and Datalake. The ILM (Information Lifecycle Management) manages Storage Class (AWS S3 and GLACIER) OnLine, NearLine Single Copy, and NearLine Dual Copy.

The functional view shows three tiers (Perf/HPC storage, Online storage, and Nearline storage). Richard displayed a diagram describing the framework.

Doug asked if the datalake has a catalog? Richard replied that it has two.

#### [DLR](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_15.00_DLR-ArchiveTechnologyAndTrends.pptx)

Katrin Molch (DLR) gave a report on current archiving technology at DLR. She reported that the hierarchical Storage Management (HSM) concept still holds; 1st data copy migrating to disk-based archive; 2nd copy on tape.

Online data holdings on exploitation platforms change role of archive from being the initial contact point in EO data flows and primary data source for users to being the “vault.” This means fewer deliveries from the archive directly to the users due to efficient upload and reload of large data volumes onto platforms and ingestion flow via HPC.

Fast transfer of mass data from archive to online storage on platforms require: upload, reload, disaster recovery.

There is renewed consideration of whether to archive intermediate and higher-level products.

Across-agency coordination and collaboration for large archive data holdings of common interest (e.g., Sentinel user level data) would seem useful.

Katrin described the archiving policy and content, and the archived data format and packaging: The data model for ingestion into data management database is specifically designed for each new collection. A simple standardized data model (like ESA EO-SIP) is also used. Metadata and quicklook components used for populating catalog. For use on platforms (internal, external), either file based or processed to ARD in cloud-optimized formats for direct use are available.

Current challenges and needs are:

* Managing multiple data copies, versions, and formats across LTA and various online storage systems (terabyte, IDA, Geoservice) 🡪 internal inventory at collection (/layer) level
* Tracking provenance; making the information available to users
* Managing multiple collection and product catalogs (ISO, EOWEB-PL, STAC) exposed internally/externally via different interfaces 🡪 “Master catalog” envisaged for new Online Data Management System (ODMS)
* Large and growing datasets of global interest, e.g., Sentinel user level data
  + Coordination of archive data holdings across agencies would seem useful
  + Can redundancy of curated “authentic copies” be achieved across agencies - including consistency checks, defined interfaces?

Discussion points:

Katrin suggested a future topic for discussion on whether redundancy can be achieved across agencies instead of maintaining redundancy within each agency.

Mirko noted that ESA is sharing several of DLR’s challenges, given the objective to try to minimize data to archive without losing anything.

Andy raised the need for decisions on when to use HPC and when to use Cloud.

Katrin noted that they have efficient interfaces should Cloud be necessary, but it is not considered part of their infrastructure, and they have no plans for using the Cloud for archiving.

Mirko commented that there is a mixture depending on the mission at ESA.

Andy agreed that having mirror sites for the redundancy is desirable.

Tom asked if multiple copies of datasets are needed instead of a single a copy for users to access? Doug replied that there is a use case for having multiple copies.

#### [ESA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_15.15_ESA%20Archive%20Technology%20and%20Future%20Trends.pdf)

Daniele Iozzino (RHEA for ESA) gave a presentation on ESA EO archive technology. He began with a listing of the ESA EO archives, including the Copernicus Programme archives. ESA and third-party active mission data are circulated by the Preservation Element Front-End. Once the data has been ingested and validated, confirmation reports are sent to the Mission Payload Data Ground Segment via standard network transfer protocol. Bulk dissemination of data coming from reprocessing campaigns is circulated on storage devices. The ESA EO Master Archive includes PDGS Data Archival, Management and Processing Services (DAMPS). ESA has outsourced EO data archiving activities to a single provider with the goal to benefit from economies of scale and standardization of data archival and delivery processes and interfaces.

In order to guarantee the data safety, the archive is distributed between two locations sited more than 200km apart; a master archive in Luxembourg and an archive back up in Sophia Antipolis (France). One of the most important activities performed by the Master Archive service is the data quality check, which is performed not in terms of scientific content but in terms of reliability of the process during data transfer.

The Space Data Preservation Archive is a data archive deployed at ESA/ESRIN premises to allow seamless archiving and extraction of data. The archive is the front-end of several long-term archiving services.

Daniele described the archive infrastructure, where data eventually written on magnetic tapes. He noted that tapes have several challenges, including tape quality, cost, and availability of vendor solutions.

Daniele described the evolution of the archive from 2014 to 2021, from a technical standpoint. He suggested as possible future use cases Object Storage as a possible replacement for the disaster recovery library where data is already unstructured and metadata driven and limitations of OS are like those of the Robotic Libraries.

Upcoming Earth Explorer missions Biomass, EarthCARE and Flex are being assessed and consequently, the archive is being benchmarked and fine-tuned to assure seamless integration with the increased amount of data produced. The expected increase of data produced in 2023 of is 5 PB and 10 PB in 2024. The industry has been engaged to suggest a roadmap to keep the archive on par with technology evolution and allow seamless archiving. Possible solutions are already being assessed.

Katrin mentioned that a presentation on the roadmap of the evolution of the ESA archive would be very valuable.

# Data DISCOVERY and ACCESS



## [International Directory Network (IDN) Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_16.00_IDN_CWIC_Report.pptx)

Michael Morahan (NASA) gave a report on the International Directory Network (IDN). He described the IDN search portal and the CWIC portal and displayed IDN/CWIC metrics, noting that the IDN offers FedEO and granules. Michael listed recent CEOS/IDN/CWIC collaboration activities (ChinaGEOSS, NRSCC, ESA/FedEO, ISRO, NOAA/NCEI).

Michael discussed the OSDD registration for RelatedURL metadata and the RelatedURL OSDD implementation for IDN granule search, giving several examples.

The IDN continues to encourage providers to add license/data usage policy information, as a link or text, to their datasets in the CMR for the IDN.

Michael described recent UMM Updates and GCMD Keyword releases. He concluded with usage metrics of the IDN homepage and the IDN search portal. He asked the participants if they thought the IDN homepage is it still needed?

Esther asked what the governance is on the services and tools. Michael replied that they are accepted from trusted agencies. It was noted that there is a high risk to expose services and tools. Valerie added that a service/tool should have a dataset to link it to, otherwise it will not be discoverable. As an initial policy the IDN team goes through all the tools and services to see if they have had updates, and if they have not, they reach out to owner. This manual approach is not expected to outpace resources. Esther commented that for Jupyter Notebooks this would not scale up – the list could become quite large. Further discussion on recommendations would be useful, harmonized with what is being done in the IDN.

Damiano noted that the discovery of products and collections is mature, and there is a need to move into the discovery of services, and the resources. Eventually this will move to more usage of those tools.

## [Federated Earth Observations (FedEO) Update](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_16.15_FedEO_Update.pptx)

Yves Coene (Spacebel for ESA) gave an update on Federated Earth Observations (FedEO). He discussed the FedEO Metadata Mediator, metrics evolution, DIF-10 export improvements and software changes, which are now visible in the IDN. DOI encoding is still to be improved.

Integration activities of ESA CCI have completed. Ongoing are the JAXA G-PORTAL (in the FedEO validation/integration environment). Additional improvements to be agreed with JAXA include:

Improve eop:EarthObservation information available in granule metadata, in particular property values (e.g., processing Level) to allow validation.

Additional (standard) queryables for granule search. E.g., gpc:polarization, eo:orbitDirection, searchTerms.

Update online resource function codes in ISO collection metadata to have access in FedEO client and Atom response.

Requests without parameters used for obtaining metrics (asset pages), often fail (time-out).

Online documentation is being prepared as part of EOCAT Evolution. This is implemented with Jupyter Notebooks and JupyterBook and includes Colab and Binder access. Documentation will be available at <https://github.com/eovoc/eo-books>.

EOVOC, reported at WGISS-54, status: Transfer to Operations of project results is being planned. Results available on public GitHub repository.

Makoto thanked the ESA team for their efforts for the G-Portal. JAXA is now preparing DOIs to reflect into the catalogue and then will request the FedEO connection.

Damiano commented on a possible evolution where the usage of the system can be implied/embedded with navigation directly to the page without going through the interface.

## [OGC Testbed-18: Secure, Asynchronous Catalogues](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_16.30_OGC_Testbed-18.pptx)

Yves Coene (Spacebel for ESA) and Liping Di (GMU for NASA) gave a presentation on the OGC Testbed-18, a yearly OGC Innovation Program initiative; one of the activities is Secure Asynchronous Catalogs.

The Secure Asynchronous Catalogs task objectives are to address research questions:

Can OGC API-Records support classical discovery workflows (OGC CSW/ISO 19115)

Establish asynchronous communication (subscribe to search results)

Apply “Data Centric Security” (DCS)

EO Use Cases:

Asynchronous Communication: Support (catalogue) users who run the same query repeatedly: get notified when additional results become available or resources change. Example: images from X over Paris with cloud-cover less than Y. Support (incremental) harvesters. Notify when new records become available or are updated corresponding to search criteria.

Data Centric Security: Exchange encrypted/signed metadata records/results making them unreadable when not authorized (e.g., confidential, paying, DRM, …)

Yves described the discovery scenario, which includes Data Centric Security, an approach to security that emphasizes the dependability of the data rather than the security of the networks, servers, or applications. This was explored in OGC Testbeds 15, 16, 17; for Testbed-18 the task is to apply it to metadata exchanged between Client/Subscriber, Catalog Server and Publisher using encryption and signatures.

Yves described DCS security. There are two alternative approaches, which heavily rely on existing JSON technology supported by open-source libraries:

(1) Client asks server to encrypt responses with a key of his (server) choice. Server passes info (kid) in response for client to find and access key via a shared KMS (key Management System) accessible via a RESTful API (See OGC 22-014).

(2) Client gives server the (public) encryption key to use. Client decodes response with his private key. No KMS is involved.

Yves described asynchronous communication, subscriptions, and several implementations (OGC CSW, OGC API-Records, Client – React, Client – Notebook).

Conclusion and next steps: Secure Asynchronous Catalog task in TestBed-18. Solutions proposed for subscriptions, data centric security and use of ISO workflows in a JSON environment (API-Records). Complementary to discovery scenarios considered in WGISS (IDN, CMR, FedEO). All information will become available in Engineering Reports OGC 22-018 and OGC 22-014 (ERs).

FedEO ISO19139 collection and granule metadata used for testing.

DCS and asynchronous solutions apply equally well to OpenSearch or other OGC (JSON) API.

Doug asked how to tackle the problem of using someone else’s API? The response was taken offline.

## [Service Metadata and Discovery Best Practices](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/4-Tuesday/2022.10.04_16.55_Service_Discovery_Best_Practices.pptx)

Yves Coene (Spacebel for ESA) gave a report on Service Metadata and Discovery Best Practices. He reminded that at WGISS-52 there was a request for collaboration to draft a “CEOS Best Practices for Service Metadata and Discovery” for elaboration within System Level Team (SLT). Collaboration within the SLT has been ongoing since then.

Extension of the approach for collection and granule discovery (CEOS OpenSearch Best Practices) with intention to share “services” (incl. tools, applications). Yves gave an outline of the documents:

* §2 - Objectives and needs
  + Use cases for service discovery
  + Detailed scenarios
* §3 - Best Practices and recommendations
  + Service metadata model (minimal)
  + Metadata encoding
  + Controlled vocabularies
  + Service discovery interface
* §4 - Description of current implementations
* Annex A: Service and Tool Metadata Elements
* Annex B: Best Practices applied to Encoding

Progress and current status:

October ‘21: Request for collaboration at WGISS-52

November ‘21: Initial draft available.

March ‘22: Update resolving review comments from SLT group on initial draft (D7).

22 March ‘22: Presentation of document at WGISS-53.

April ‘22: Updated draft (D8) with additional SLT inputs.

May ‘22: Updated draft (D9) for review by wider audience – incl. coupled resources.

June ‘22: Presentation of document at Data Access and Preservation DAP#10 WG Meeting.

July ‘22: Request for feedback to WGISS

Sept ‘22: Reminder for feedback to WGISS

Comments were received on ARD and learning resources. Next steps are to include “coupled resource” recommendations (“Under Review”) as no objections raised, and prepare final version.

Damiano urged the team to carry on with success.

# Data INTEROPERABILITY and USE



## EO Interoperability for Services

Andrew Mitchell (NASA) introduced the question of whether WGISS agencies would benefit from developing guidelines/best practices for reducing the burden on data repositories of needing to archive multiple copies of their data.

Discussion points:

* Are there new technologies or processes that can aid in the bulk sharing of data between agencies?
* Should WGISS rely more on pointing to other agency data directories/catalogs (via indexing) instead of data exchange?
* Can WGISS provide guidelines for how to store data in commercial clouds in order to maximize data use? As an example, WGISS could recommend that archives are stored in the same cloud region (e.g., AWS -West Oregon) to prevent data transfer costs to the user. NOAA (Ken) mentioned programs where the commercial providers have equally archived popular datasets without charging their agency. This has allowed broader access to datasets.
* What can/should WGISS do to prepare for the deluge of data copies available that are not from the certified data source? Should agencies certify entities to distribute copies of their data?
* Should/can WGISS provide a study on when to use high performance computing versus cloud computing? ISRO has performed a similar study.
* Is having data available on various platforms to cut down the redundancy an option?
* Is it advisable to make visible to the user where the data is stored to aid the processing.
* Discussion of ways of reducing the burden on the user.
* Modelling under HPC.
* HPC on premises with data residing on the cloud and user pull-down small amounts as needed.
* The data in multiple clouds is a challenge; need ways to authenticate the data. Encourage pointing to the data instead of multiple archives. Is there a way to encourage cloud vendors to take on some of the responsibility to maintain an authentic copy; it is to their benefit.
* It is very difficult to convince users to migrate to cloud until they encounter limitations of on premise HPC. User must go to hybrid approach or move to cloud. ISRO Performed a study of getting those facilities to the users using cloud. The study will be shared with WGISS when it is ready.
* The reluctance to use cloud will change with time. But for underserved communities that model (not cloud) will always need to be maintained.
* NOAA has been working on agreements with cloud providers to host data while giving free download to users. Providers and agencies are finding creative ways to manage the costs.

## Session on EAIL: Use-cases and Vision

### [CEOS Coastal Observations, Applications, Services, and Tools (CEOS COAST)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_09.40_COAST%20EAIL.pptx)

Paul DiGiacomo (NOAA) and Merrie Beth Neely (NOAA) gave a presentation on CEOS Coastal Observations, Applications, Services, and Tools (CEOS COAST). This is a Coastal Zone (land-sea-people interface) Focused Team to leverage global Earth observation data (satellite + in situ) to tackle coastal problems affecting society - Coverage | Frequency | Resolution.

EAIL has enabled COAST product developers to readily transfer technology and leverage processing efficiencies. COAST has trained both the NOAA and ISRO teams in use and are poised to embrace EAIL more fully. One clear message is that as users see high resolution data, they want more, especially higher spatial for coast users who engage in co-design and co-development making sure the data (including in-situ data) is something they need.

CEOS COAST has identified two Pilot Projects to develop coastal products, services, and tools. They have identified pilot areas that are continental, and small island nations. COAST is focusing on sea to land impacts in the areas of biological, water quality, submerged vegetation, habitat mapping, coastal eutrophication, turbidity, bathymetry, flooding, and modeling.

The EAIL helps with integration, informatics, providing a laboratory of common framework to work and trade best practices and interoperability. They hope the EAIL to grow and expand.

Paul discussed CEOS COAST trends in parameters. The more that can be done with the data to help the users… derived products to information and knowledge. The data itself is too much. Trying to do curation, extract actionable information, but users can still use these environments to do it themselves.

CEOS-COAST Application Knowledge Hub is meant to bring together (curating) fit for purpose products for users in the coastal zones, using environmental and socio-economic data together, harnessing satellite data for environmental justice.

Paul listed several upcoming product development areas and pilot regions. The beta release and expansion of CEOS COAST Application Knowledge Hub will be soon. The team expressed the need to ingest and utilize data/products from additional CEOS missions, and more emphasis on using the CEOS EAIL which will require further sustainment and development support to expand capabilities and data offerings.

CEOS COAST is also looking for interested CEOS agencies to provide active leadership and support in these new and expanding areas. They need more data and products so they can provide these fit for purpose products. They also hope to see a huge expansion of the EAIL but will need support and engagement from CEOS agencies.

Paul concluded thanking CEOS for the vision of the EAIL, a wonderful resource that brings out the best of CEOS to enable getting things done.

### [CSIRO](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_10.00_EAIL_CSIRO.pptx)

Jonathan Hodge (CSIRO) gave a presentation of the EAIL from the point of view of CSIRO Chile, where it has been operating for the last two years. The EAIL is built using the Open Data Cube software and CSIRO’s Earth Analytics, Science, and Innovation platform.

The EAIL is collection of components that allow analysis and visualization, and scalable work flows. The information can be put into a format that is useful to analyze the data. Significant data updates have occurred over the last 6 months including increased coverage and new products. There are a range of potential data sources, services and applications which could be added to EAIL if there is interest.

Numerous other data sources either under development or in operation in related projects which could benefit EAIL, including: MODIS (numerous products), Sentinel 5P, GEDI. Related projects (e.g., Aquawatch Australia) are improving integration with time-series data sources, enabling improved connection with sensor and cal/val data.

Potential application capabilities include:

* OWS service layer ready for implementation: WMS, WMTS and WCS services, exposes data cube layers to the outside world, supports multiple styling options, band combinations and index generation
* Terriamap web visualization tool: Connects directly to WMS services above, supports over 30 data and services type including time-series data and processing services (WPS)

New and potential analysis capabilities are GPU processing with AWS GPU nodes, additional scientific programming options with R, and new machine learning capabilities.

Current issues and challenges:

* Detecting and resolving issues in third-party catalogues. Some data catalogues are a “best effort” initiative which inevitably can contain small but significant errors. New workflows are in development to identify and respond to these errors which should be fixed at source. The current issues affect Sentinel 2 and Landsat 8/9.
* User support and training: New ML and GPU capabilities bring huge opportunities but will require a deeper level of support and discussion.
* Financing and ongoing support.

Jonathan concluded with potential improvements and additions, and listed future use cases.

Paul noted that AI and ML approaches to extract more features from the large amount of data will be significant.

Jonathan added that there are issues in the public STAC catalogues that need to be resolved. As an example, there are STAC records which are pointing at data on the AWS cloud that no longer exists, so the STAC records are wrong; there are also challenges with user training and support - GPU and ML work is going to require sharing the work more.

## STAC

### [Federated Discovery using STAC](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_15.30_Federated%20Discovery%20using%20STAC.pptx)

Doug Newman (NASA), Yves Coene (ESA, Spacebel) gave a presentation on a STAC API for federated search.

STAC API solves use cases outside of the federated use case (cloud-friendly discovery, for example) and has all the features that made Open Search attractive for the implementation of a federated search architecture.

Some CWIC providers have developed their own STAC APIs ([NASA](https://cmr.earthdata.nasa.gov/stac), [INPE](https://stacindex.org/catalogs/cbers)), and there is a vibrant community and tooling (all open-source). STAC is specific to the Earth science data domain.

For federated search a collection search endpoint definition, a collection search parameter, and an item search endpoint definition are needed, as is a programmatic linkage between distinct catalogs.

The proposed API extension would allow links to Item Search endpoints from extra-domain catalogs in a catalog. The STAC API currently assumes the user knows the collection(s) the user requires and concentrates on item-specific dimensions. To be a federated solution, the search at the collection level needs to be opened. OGC has a proposed API containing collection-level search support that suits the need. STAC is considering using that to solve the same problem. This raises the question of whether WGISS and agencies should remove that from the extension and assume that STAC will implement something similar?

Next steps are to implement federated extension in CMR with Pathfinder CWIC provider, and decide on waiting for STAC to integrate with OGC with respect to collection search, or forge ahead. Doug recommended forging ahead, since it is unlikely OGC will pivot, but they will take some time. Michael cautioned that CMR STAC does not address all the requirements for federated CWIC.

### [Access Patterns and STAC Implementation at Bhoonidhi](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_15.40_Access%20Patterns%20and%20STAC%20Implementation%20at%20Bhoonidhi.pptx)

T. Sai Kalpana (ISRO) and K. Radha Krishna (ISRO) gave a presentation on access patterns and STAC on Bhoonidhi at ISRO.

ISRO’s huge repository of data can be found in Bhoonidhi (EO missions) and MOSDAC (meteorology, oceanography, and tropical water cycles). Data access of Bhoonidhi is open with various mechanism for dissemination. For MOSDAC, ISRO has decided to move to STAC.

They described the requirements, the POSTGIS STAC support, the STAC-FAST-API, and the Bhoonidhi STAC architecture, STAC Browser, and API gateway features.

The current roadmap is:

* Load testing and integration testing in works
* Open API SDK generator being explored for python, java
* Integration of STAC with Bhoonidhi code lab in progress
* Estimated open implementation by Q1 2023

Doug thanked them for helping with the federated STAC proposal.

Krishna clarified that optimization is built into the system based on access patterns.

### [STAC Best Practice](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_16.10_STAC%20Best%20Practices.pptx)

Doug Newman (NASA) discussed best practices for STAC. The top six with respect to WGISS and the CWIC are:

Practice 1: Organizing your catalog hierarchy.

Practice 2: Enable CORS

Practice 3: Be compatible with STAC browser

Practice 4: Asset-oriented catalogs

Practice 5: Catalog decisions

Practice 6: Searchable Identifiers

Yves asked if the STAC browser combination is used with dynamic catalogs; Doug replied that the CMR example works well. User interfaces are helping STAC browser mature; looking to expand in the areas of drill down.

## Discussion

Makoto Natsuisaka (JAXA), Tom Sohre (USGS) led a discussion on WGISS participation in and endorsement of the CEOS Interoperability Framework. The ToR is being drafted and will be presented to the CEOS Plenary. Once approved at CEOS Plenary, WGISS will be asked to nominate participants in the task group, and WGISS may be assigned actions. Makoto, as WGISS Chair, would like to participate.

Discussion points:

Peter Strobl suggested standardizing ARD though the targets are very broad. If CEOS ARD becomes standard, it will need to match resources. The concept is important, but implementation, who will do the implementation?

Makoto said interoperability is also very broad. But the category of Data interoperability and federation are WGISS issues. Richard and Mirko agreed that WGISS needs to participate as coordinator for the technical aspects as this is the core business of WGISS; they also agreed to participate.

Tom reminded that there is no leader for the WGISS Data Use and Interoperability Interest Group. He invited volunteers, and added the option that the participation occurs across all the groups.

**Action WGISS-54-12: Makoto Natsuisaka to share CEOS Interoperability Framework to WGISS, and call for participants.**

# TECHNOLOGY EXPLORATION



## Jupyter Notebooks

### [Update on Jupyter Notebooks Day](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_09.40_Jupyter_Notebooks_Day.pptx)

Esther Conway (UKSA) gave an update on the upcoming CEOS Jupyter Notebooks Day – Supporting Global Best Practice and Training, Friday October 21, 2022. As background, she noted that the CEOS WGISS and WGCapD ran a joint webinar on Jupyter Notebooks for Capacity Development whose aim was to introduce space agencies and environmental organisations worldwide to Jupyter Notebooks and take a tour of emerging services from CEOS Agencies and their applications. How Jupyter Notebooks can be used to support capacity development and the exploitation of Earth Observation data by a broad range of users was illustrated. The registrants were surveyed and they responded with two key requirements:

* The development of a Jupyter Notebooks Best Practice
* Jupyter Notebooks-based training

The Jupyter Notebooks Day event will be a working meeting to facilitate solutions for those needs. The agenda includes the following:

* Round Table asking specific technical questions
* Review of Best Practice
* Discussion of Training Capabilities and Requirements

Esther listed the Round Table Technical Questions and gave an outline for the Best Practice document.

The discussion of training requirements and capabilities will include:

* How CEOS can support EO training using notebooks.
* Who would be willing to contribute (e.g., provide notebooks, platforms/services, trainers, and help organize or support participants).
* What should a first event should look like?

Esther listed the core and extended group of participants, and gave the link and logistics for the webinar. She will send out the invitation to WGISS-all and other targeted participants, and Kent Ross agreed that the event would be more successful by staying focused on CEOS objectives.

### [Jupyter Notebooks Best Practice Document](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_10.00_Jupyter_Best_Practice.pptx)

Esther Conway (UKSA) discussed the draft Jupyter Notebooks Best Practice document. She noted that EUMETSAT has joined the core group since they have also been working on this and have examples and participated in the Living Planet Symposium. Esther showed the document outline, and listed the target communities.

Who will benefit from a Best Practice: Data producers, authors of Jupyter Notebooks, providers of EO data training, users of EO data, EO data archivers, and providers of Data Analysis Infrastructure

The objectives and needs, issues to be addressed, of the document include:

1. Clear understanding the purpose of a notebook and deciding if it is a reusable asset.  Successfully conveying how and why a notebook should be used by its intended community.
2. Encourage the development of good workflow and structure within notebooks along with quality documentation. Support their reuse and adaptation by new users.
3. Support discovery of relevant notebooks in terms of dataset, domain, application/function, and skill level.
4. Delaying technical obsolescence and ensuring longevity of relevant notebooks.
5. Maintaining the quality of archive by timely removal of redundant notebooks.
6. Supporting interoperability of notebooks on different platforms.
7. Lowering the barrier for EO data exploitation and raising technical skill level.

The best practice will encourage the development of good workflow and structure within notebooks along with quality documentation, and supporting reuse and adaptation by new users. Esther detailed Notebook description and function, technical dependencies and virtual environments, citation of input data and data access, association with archived data, version control, preservation, archival, software licensing, and advanced topics (interoperability and reuse on alternate platforms, creating other deployments, and incorporation with data cubes).

As future topics she suggested that a registry to help new users navigate the landscape is needed and to help the full range of agencies and developing countries engage.

Discussion points raised:

* Focus on usage differences; initiate the licensing discussion.
* Categorization of notebooks.
* Addressing maintenance of a Jupyter Notebook, with proper declaration of technical dependencies, and not allowing accumulation of obsolete notebooks.

### [WGCapD Updates and Webinar Support](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_10.20_WGCapD_Status.pptx)

Kenton Ross (NASA) discussed upcoming WGCapD activities and listed the potential WGISS-WGCapD Collaboration Points for 2023-2025 Workplan Deliverables:

* Relevancy Ranking of Data Search Results
* Data Cubes for Large Scale Data Analytics
* The Burgeoning Role of Python for EO Data Analysis
* Explore Capacity Development with the EAIL
* Joint Support for CEOS ARD
* Jupyter Notebook Best Practice Documentation/Training

Kenton also listed relevant CEOS Workplan deliverables for WGISS and WGCapD engagement and collaboration.

Kenton agreed to connect Valerie Dixon to the MLHub team for collaboration.

## Machine Learning, Data Learning, Artificial Intelligence: Session Objectives

### [NOAA Center for Artificial Intelligence](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_10.30_NCAI%20for%20CEOS%20WGISS.pdf)

Robert Redmon (NOAA) gave a presentation on NOAA’s progress toward an AI-ready agency. To this purpose NOAA created the Center for Artificial Intelligence. This center will be tasked to train on building trustworthy and equitable AI applications for Digital Twin Earth systems fueling economic and equitable climate policies and for automating and augmenting stewardship of national critical ecosystem resources and environmental observations.

The recommendations are to:

* Develop AI training standards centered around interactive training tools as “Learning Journeys” demonstrating Lessons Learned. This ensures AI applications are Trustworthy, maximizes technology transfer and directly benefits Research to Applications and Sustainment.
* Develop AI-ready data standards for open environmental data. This is the future of data stewardship, building upon Findable, Accessible, Interoperable and Reusable (FAIR principles) and Analysis Ready partnerships.

For AI-ready data what is needed is community-driven definition of AI-readiness requirements - assessment tools for data providers and stewards and users - representation of readiness level conditioned on use / application - feedback and iteration to improve the standard - a formal standard. This will benefit from their collaboration with ESIP AI-Ready Data.

The NCAI will begin with a pilot project: Create tropical cyclone model training dataset.

To address needs, resource creation should be prioritized to convert NOAA AI success stories into interactive training materials in a sandbox computing environment that allows the workforce to apply learning outcomes to support NOAA’s mission via the AI Strategic Plan.

NCAI GitHub will be established in 2023. They would like to collaborate with the WGISS Jupyter Notebook team.

### [Enhancing Earth Observation Applications Using AI/ML Techniques](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_10.50_ISRO%20Enhancing%20EO%20Applications%20Using%20AI.pdf)

Nitant Dube (ISRO) gave a presentation on enhancing EO applications using AI/ML techniques. Nitant noted that the Gartner report suggests maximum utilization AI/ML in the field of forecast and prediction for Earth Science Studies.

Activities in Earth Sciences that would benefit from AI/ML include forecasting, nowcasting, downscaling, retrieval, decision support, modeling, and feature engineering for generation of meaningful information from the data. Disciplines that would benefit from AI/ML include urban planning and monitoring, agriculture, weather and climate, renewable energy, disaster mitigation and recovery, hydrology, ocean and marine ecosystem, and environment.

Nitant listed several possible use cases in EO applications. Several have achieved some maturity, and others not, perhaps because datasets are not available.

* AI based model for detection of cyclone in an image and its intensity estimation with moderate success.
* Machine Learning based Soil Moisture Retrieval over Indian Cropland using space-borne GNSS-R mission.
* Classification of Hyperspectral data (ROSIS) has shown a good accuracy of 96.93 %.
* Deep learning techniques used for Soil studies (un-mixing of red and black soil) for airborne hyperspectral data.
* AI based forecasting of NDVI and wetland monitoring (this is being used operationally).
* Automatic Building Foot-print extraction and generation of 3D city model as part of Smart City Project using Deep learning Techniques.
* DL plugin tools developed for AHYASS software for Hyperspectral data analysis.

AI/ML helps in filling the Gap/ Enhancing the current understanding of science to help Earth Observation Applications.

Discussion points:

* Use of AI/ML in EO applications: this information can be used for strengthening the BP activities of WGISS.
* Tagging those that have been retrieved using AI techniques.
* Tagging the products at another level, such as AI model or numerical model, useful information to the user.
* Hyperspectral image classification.

### [Machine Learning and Graph](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_11.30_Machine%20Learning%20and%20Graph.pptx)

Doug Newman (NASA) gave a presentation on machine learning and graph. Enabling Machine Learning requires training data sets are an essential tool for Machine Learning and Artificial Intelligence (AI/ML) methods… it is critical that they are archived and cataloged… should be indexed in NASA’s Common Metadata Repository (CMR) and made discoverable alongside data in CMR.

Improving cloud-native data access can be achieved using metadata to describe access (region, bucket, S3 URL), investigating balance between access patterns and metrics based on data access experiences from the community.

Leveraging machine learning can occur using ML and Natural Language Parsing techniques to map collections to a set of GCMD keywords based on the collection description.

Enabling Knowledge Graphs: The CMR will expose a low-level graph traversal API for their collection-centric graph database using standards-based APIs, with the intention to fuel the development of novel derived knowledge extraction techniques in the community. CMR exposes a GraphQL API that exposes a schema-based query mechanism for relationships between collections, services, variables, etc.

Leveraging knowledge graphs can happen with Earthdata Search and ESIP’s usage-based discovery.

The Earth Science Data System Working Groups (ESDSWG) is a NASA organization that focuses on the exploration and development of recommendations derived from pertinent community insights of NASA's heterogeneous and distributed Earth science data systems. The Machine Learning Capacity Development Working Group aims to develop training methods to improve the ability of domain scientists and/or information learn and apply machine learning techniques.

Activities:

* Task 1 - Machine Learning on AWS
* Task 2 - Survey on Earth AI Applications
* Task 3 - Machine Learning Tutorials
* Task 4 - Machine Learning on NASA’s Earthdata Cloud

ML cluster activities involve sharing experiences and resources on machine learning and its applications in the Earth and space sciences, and developing training materials to share with the broader ESIP community.

Discovery cluster activities explore usage-based discovery, including relationships between datasets and types of usage, from publications to applications to policy-making.

Discussion points:

* Anyone participate in the ESDSWG - to be confirmed
* Survey on Earth AI Applications to be shared with WGISS
* The GCMD keyword recommender has an accuracy of 80%. Not sure yet if the accuracy is increasing over time; the feedback functionality has been implemented, and data is slowly being introduced to help with curation.

### [USGS AI/ML Use Cases](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_12.00_USGS%20AIML.pdf)

Tom Sohre (USGS) gave a presentation on AI/ML workflows with EO data at USGS.

USGS EROS AI/ML workflows generally make use of similar tools and data for informing understanding of Earth surface and subsurface conditions, changes, and drivers. Software libraries used include: STAC, DASK, Xarray, GDAL, Pytorch, and Tensorflow. Compute platforms include Amazon Web Services, Microsoft Azure, Google Cloud Platform, and High-performance computers (HPCs). Data used is active and passive remote sensing imagery (e.g., Landsat Collection 2, Harmonized Landsat and Sentinel-2, Sentinel-1 and 3, Planet, MAXAR, IceSAT-2).

Targets of interest are:

* Land cover conditions (e.g., thematic, fractional cover), including Impervious Surface mapping for National Land Cover.
* Land surface phenology, looking at invasive grasses, several different datasets using neural networks
* Species and lifeform level mapping.
* Neural Network model used to predict the cover of several rangeland components simultaneously. Test accuracy demonstrated a ~9% increase in accuracy moving from regression trees to neural networks.
* High resolution UNET tree cover.
* Surface water conditions.

Scalable AI/ML workflows with EO data challenges and opportunities:

* Filling gaps between software libraries commonly used in geoscience data analysis (e.g., Xarray, Dask) and libraries commonly used for deep learning (e.g., TensorFlow, PyTorch).
* Lower barriers to entry and improve modelling tasks. Building catalogs/libraries of deep learning models, EO training/testing data, and reproducible workflows (e.g., Jupyter Notebooks).

Tom concluded saying that most of their initiatives are project-based and in the exploratory phase, and not yet a large-scale effort.

### [New AI/ML Keywords Introduced to GCMD](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_12.15_AI-ML%20GCMD%20Keywords_2.pptx)

Valerie Dixon (NASA) discussed AI/ML GCMD Keywords. She began with a brief description of the GCMD, which now includes Earth science services categories. The keywords in the Earth science services category are:

* DATA ANALYSIS AND VISUALIZATION
* DATA MANAGEMENT/DATA HANDLING
* EDUCATION/OUTREACH
* ENVIRONMENTAL ADVISORIES
* HAZARDS MANAGEMENT
* MACHINE LEARNING TRAINING DATA
* METADATA HANDLING
* MODELS, MACHINE LEARNING MODELS
* REFERENCE AND INFORMATION SERVICES
* WEB SERVICES

Earth Science Services include machine learning training data for running a machine learning model (labels and source). Valerie also listed several machine learning models.

Best practices are still being defined and refined, but considerations include:

* Apply ML Training Data keywords only to ML training data
* Apply ML Model keywords only to ML model data

Although there are use cases where it might be useful to use both model and training data keywords (e.g., this ML model was built using these training data types, or this training data sat is useful for these ML model types), these kinds of cross-associations are better done in other ways: Collection/Service associations, to be determined CMR Graph relationships, Abstract or Purpose fields, etc.

Keywords to support AI/ML are gradually being ingested into the MLHub. Once it is there, they will be in the IDN for anyone to find.

### [Introducing White Paper for AI/ML Outline](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_13.30_TechEXPO_MLDLAI_WhitePaper.pdf)

Yousuke Ikehata (JAXA) discussed the work on the white paper for AI/ML. He began with the background and scope:

* This white paper shows a use case about Deep Learning (DL), Machine Learning (ML) and Artificial Intelligence (AI), using EO data.
* Some CEOS agencies provide DL/ML/AI data and platforms and introduce them at WGISS meetings.
* There are not any summarized documents.
* This document can lead data users to data and platforms, and data providers can know where to upload or what cloud is proven for DL/ML/AI datasets.

The target audience is data users such as students and data scientists, and data providers, agencies with plans to provide datasets for AI/ML. The outline of the white paper is:

1. Introduction

2. Basis of AI/ML for earth observation domain: Event analysis, learning based analysis, simplify data analysis like image processing.

3. AI/ML use cases in CEOS.

* Categorize AI/ML application
* Estimation
* Techniques that can rediscover EO data value
* Show experimental use case: –UK/PML –ESIP –NASA ESDSWG Task 4 • ML training data and NASA Earthdata Cloud –NOAA

4.Share platforms and data: Show/List platforms and data for AI.

5. Clarify the difference between general AI/ML and EO’s AI/ML.

Ken Casey noted that the NOAA Center for AI has done work related to this, but is NOAA-focused. This paper could be valuable and complementary to what NOAA has done. He remarked that one thing that might be useful is the [NOAA AI strategy document](https://sciencecouncil.noaa.gov/Portals/0/2020%20AI%20Strategy.pdf). It may have some utility in drafting this WP.

**Action WGISS-54-13: Technology Exploration Interest Group to add the production and publication of the AI/ML White Paper to the CEOS Work Plan. Due by WGISS-57 (April 2024)**

## [Federation](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_14.45_Federation.pptx)

Karim Ramage (CNRS) and Richard Moreno (CNES) gave a presentation on different aspects of federation. The community faces challenges which require hybridization of a huge variety of data, stored and exploited in a huge variety of platforms. This may require that organisations and platforms cooperate (exchange data, processing) with a mind of good efficiency (e.g., no data download or duplication, prefer big infrastructure).

Examples of Federation are:

* EOSC – A web of FAIR data and services for science
* Destination Earth – develop a very accurate digital model of the Earth
* DestinE Data Lake (EUMETSAT)
* Data Terra – French Research Infrastructure Data and Services of the Earth System
* GAIA-X – A Federated and Secure Data Infrastructure. For example, European clouds federation to foster data usages

Different aspects of federations derived from the GAIA-X model:

* Iaas – infrastructures: cloud, HPC, Edge instances, Network. Compliance, portability interoperability and interconnectivity are all considered.
* Daas – EO data and services. Types of data: data access and services, space data, data Terra RI, Copernicus open hub. Compliance, portability interoperability and interconnectivity are all considered.
* Paas – Platforms. Types; mix data, services and infrastructure plus AI, DIAS, GeoDataHub, Earthdata, and Google Earth engine, SentinelHub, and ESA TEP. Compliance, portability interoperability and interconnectivity are all considered.

Federation services:

* Federated catalogue and discovery portal for data and services (inc. Notebooks, Data cubes, …)
* Identity and Trust
* Sovereign Data Exchange
* Compliance (legal: regulation and policies)
* Governance
* Rules for onboarding in the federation
* Cross services

Discussion points:

* Determine what is needed for the federation.
* Propose federation as a system (technical, governance, security, architectures, standards)
* Document next steps.

**Action WGISS-54-14: Technology Exploration Interest Group to develop in-depth presentations at WGISS-55 on federation, including a call for proposals for federation. Due WGISS-55.**

# WGISS-WGCV Joint Symposium



## [WGCV Overview](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_13.30_WGCV%20Introduction.pptx)

Akihiko Kuze (JAXA), Chair of WGCV, introduced the activities of WGCV. Recent topics include:

* WGCV will nominate Cody Anderson of USGS as VICE chair 2023-2024 (WGCV-51)
* WGCV to ‘New Space’ – legacy experience, infrastructure, tools (SIT-TW)
* Action on CEOS Interoperability Framework (SIT-TW)
* WGCV will collaborate with the International Methane Emissions Observatory (IMEO)

WGCV subgroups are:

* Atmospheric Composition (ACSG)
* Infrared Visible Optical Sensors (IVOS)
* Land Product Validation (LPV)
* Microwave Sensors (MSSG)
* Synthetic Aperture Radar (SAR)
* Terrain Mapping (TMSG)

## [WGISS Overview](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_13.40_WGISS_Overview.pptx)

Makoto Natsuisaka (JAXA) introduced WGISS. He listed recent topics:

* Data Integrity and Authenticity on Cloud -> under discussions
* Data Quality Assessment and Indicators, and DMSMM Maturity Matrix -> Best Practice
* WGISS Connected Data Assets -> in operation
* STAC -> Best Practice
* Service Discovery -> Best Practice
* CEOS Common Online Dictionary -> Online Dictionary
* EAIL (Earth Analytics Interoperability Lab -> EAIL
* Jupyter Notebooks Initiative -> Best Practice and Webinar
* Machine Learning, Data Learning, Artificial Intelligence -> White Paper

Cooperation with other entities:

* CEOS ARD Oversight Group … CEOS SIT, VCs, WGs
* CEOS Ocean Coordination Group … CEOS SIT, VCs, WGs
* CEOS Interoperability Framework … LSI-VC
* CEOS Engagement with Standard Organizations … LSI-VC
* EAIL initiative … SEO, WGDisasters (Flood Pilot Project), WGCV (DEMIX), COAST, Rice Monitoring Community
* Jupyter Notebook initiative … SEO, WGCapD
* Common Maturity Metrics … WGCV, CGMS
* Common On-line Dictionary … WGCV
* AI/ML White Paper … WGDisasters
* Service Discovery … WGDisasters

## CEOS CAL/VAL Portal, ESA EO Portal, CEOS MIM Database: Status, Synergies, and Improvements

### [ESA eoPortal and CEOS MIM Database](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_13.50_eoPortal%20CEOS%20MIM%20Database.pptx)

George Dyke (Symbios for ESA) gave a presentation on the ESA eoPortal, which provides a trusted, accurate, and up-to-date gateway to knowledge and resources related to Earth Observation satellite missions. The portal facilitates explaining, demonstrating, and promoting the capabilities of space systems. The portal is comprehensive and rich with technical detail and is the core EO reference for the community.

The eoPortal has been modernized with an improved user experience, boosting engagement from the Earth observation community. The portal has increased emphasis on analytics and web performance, and is more secure. The satellite missions’ catalogue has been improved, leveraging CEOS MIM Database with improved integration for robust indexing/framework.

Future plans are to refine current integration of CEOS Database information in the EO Portal, consider how to enrich with measurements, GCOS, SDGs, disasters content from CEOS Database, and perform further development of the content.

Matthew Steventon (Symbios for ESA) gave a presentation on the CEOS MIM Database, which contains information on EO satellites (Missions, Instruments, Measurements -Datasets) based on an annual survey of CEOS member agencies. This database represents the only official consolidated statement of agency programmes and plans, and provides a community focal point for the coordination of future planning, research, and gap analyses, and providing an interface for the user community.

New developments in the MIM are gap analysis tools, the CEOS Database Toolkit - Python Notebooks, and several analyses used for internal purposes. Other developments are a pilot implementation of instrument response curves for analysis, looking to build on this capability in future based on user needs and project work. A pilot API was built 3-4 years ago and was a successful proof of concept but use cases not yet developed. Several small studies into using CEOS Database along with the COVE API showed promise.

### [CEOS CAL/VAL Portal](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_13.50_CalVal%20Portal.pptx)

Paolo Castracane (ESA) gave a presentation on the CEOS Cal/Val Portal, which serves as the main forum for exchange and information sharing for the CEOS Working Group on Calibration and Validation. The portal provides access to agreed good practices and cal/val protocols to the wider Earth Observation community within CEOS and beyond. It also connects users to reference data and networks and provides reliable, up-to-date, and user-friendly information useful for cal/val tasks, facilitating data interoperability and performance assessment through an operational CEOS coordinated and internationally harmonised cal/val infrastructure consistent with QA4EO principles.

The portal is the entry point for all the WGCV subgroups, provides access to links, news, events, documents, projects, tools, and cal/val data and campaigns.

Paolo confirmed that data access occurs using links, and a user must be registered on the portal, and is assigned a specific role (subgroup related, etc.)

### Discussion

Matthew introduced the following discussion points:

* Thoughts on improvements or additions?
* Connections with other CEOS systems?
  + Links to WGISS work?
  + API or other connectivity requirements or desirables?
  + Cal/Val Portal?
* Gap analysis tools
* Data discovery aspects
* WGISS guidance on response curves

Suggested improvements are a link between the MIM and the ECV inventory, and automation for the data providers to enter their information.

## [Status of ARD Activities](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_14.20_CEOS-ARD%20Status.pptx)

Matthew Steventon (Symbios for ESA) and Ferran Gascon (ESA for the ARD Oversight Group) gave a presentation on the status of CEOS ARD activities.

The CEOS motivation to engage in ARD activities is to facilitate data usage by reducing data processing burden on users, looking toward a future of sensor-agnostic Earth data where the uptake and impact of all data sources for societal benefit is maximized. CEOS-ARD is a key step within the ‘interoperability spectrum.’

Products assessed and compliant to CEOS-ARD specifications are known by the community to meet the high standards set by CEOS space agencies. Compliant products can be branded as such with the CEOS-ARD logo, product family specifications define parameters which facilitate interoperability. This is a subset of what is needed to achieve full interoperability.

CEOS, with its long heritage, experience and expertise has a key role to play in defining ARD for the community. The Land Surface Imaging (by LSI-VC) led the way by, and CEOS-ARD Oversight Group is now extending to other domains.

Overwhelming feedback from industry is that formal standards are needed for ARD which will facilitate the implementation of CEOS-ARD into operational workflows and to ensure products interoperability.

SIT-TW-2022-03: CEOS agencies to consider resourcing people with CEOS ARD heritage to join the OGC ARD Standards Working Group.

The CEOS-ARD OG members were listed.

## Data Quality Assessment and Indicators, and DMSMM Maturity Matrix

### [Data Quality Indicators](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_15.10_DMSMM%20Data%20Quality%20Indicators.pptx)

Philippe Goryl (ESA) gave a presentation on data quality indicators, which can be expressed in different layers of information (per pixel, per measurement, as cal/val results summary, as summary quality in the metadata, and in tools and procedures).

An important aspect of EO data products that ensures they are most easily accessible to the widest variety of users is their format. Product metadata and flags offer users important extra layers of useful descriptive information in addition to the measurements themselves that can be crucial to their analysis.

In the ideal case, an assessed mission product format should meet any appropriate CEOS Analysis Ready Data (ARD) metadata guidelines, for example CARD4L requirements in the case of surface reflectance products.

A cal/val maturity matrix provides a reader with the high-level summary of the output of a quality assessment. The matrix contains a column for each section of analysis and cells for each sub-section of analysis. Sub-section grades are indicated by the color of the respective grid cell, which are defined in the key.

The following aspects are recognized crucial for an optimal EO data quality assessment:

* The overall strategy shall be in line with the QA4EO principles
* It is optimal to have quality information per pixel and at lower processing level
* Uncertainty Characterization for EO Product shall follow standard guidelines
* Cal/val activities shall be part of the overall Satellite Mission
* Metadata shall include quality information and follow good practices and standards
* Post-launch cal/val activities shall exploit appropriate infrastructure/methods traceable to SI (e.g., FRM).
* Validation plans are implemented involving: Mission Performance Centers and Validation Teams
* State-of-the-art algorithms, tools and procedures used, easily “fit for purpose” in terms of the mission’s stated performance.

### [Data Management and Stewardship Maturity Matrix: Introduction, Background and Definitions](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_15.10_Data%20Management%20and%20Stewardship%20Maturity%20Matrix-%20introduction%20background%20and%20definitions.pptx)

Iolanda Maggio (RHEA for ESA) gave a presentation on the Data Management and Stewardship Maturity Matrix (DMSMM). DMSMM defines all activities needed to preserve and improve the information content, quality, accessibility, and usability of data and metadata. It provides a way to measure the status of the agency data stewardship processes in place, and to plan goals of data stewardship processes and projects.

Iolanda displayed the DMSMM, listing the areas, levels of maturity, and components:

Areas:

* Discoverability
* Accessibility
* Usability
* Preservation
* Curation

Levels of Maturity

* L0 Not Managed
* L1 Partially Managed
* L2 Managed
* L3 Fully Managed

Components

* Metadata for Discovery
* Online Access
* Data encoding
* Data Documentation
* Data Traceability
* Data Validation
* Data Metrology (e.g., Uncertainty)
* Data Quality Control
* Product Details
* Data Preservation
* Data Verification
* Data Processing/Reprocessing
* Persistent and Resolvable Identifier

Iolanda described the step-by-step process to evaluate each area/component, and to score the level of maturity of each area/component.

### [AVHRR DMSMM Use Case](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_15.10_DMSMM_UseCase.pptx)

Paolo Castracane (RHEA for ESA) described the [AVHRR DMSMM Use Case](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_15.10_DMSMM_UseCase.pptx). The evaluation is being performed on the AVHRR L1B Dataset using the Cal/Val Maturity Matrix (a subset of the DMSMM), and is currently in progress.

Paolo also gave a logical overview of the Maturity Matrix Tool which has three components: configuration, maturity questionnaire, and maturity matrix. He also discussed the proposed architecture.

### Discussion

This is a good example of working group cooperation, and a great achievement. The outcome will be a white paper, which is a CEOS WP deliverable for 2023.

The MMT is only for the WGCV columns of the DMSMM, but it can be configured to the other columns.

The DMSMM is helpful to find a balance between what the user wants and the number of resources the data provider will need; the analysis can be very expensive.

The evolution of the MMT can be presented to SIT, and synergies can be sought with the existing NOAA tool.

Are there plans to extend this to handle downstream products (which have a DOI referencing the primary dataset whose landing page would contain the information)?

With the recent inclusion of datasets in the MIM database, maybe the maturity matrix information could be included.

The WGISS MM is very useful and inspiring. Various groups would like to provide feedback on it – specifically the use of the term “level” (as opposed to processing level), the order of the columns, the number of columns under validation.

The Cal/Val MM has a different objective/perspective than the WGISS MM. The Cal/Val needs to assess commercial data and can compare results. A convergence between the two is being attempted.

**Action WGISS-54-15: Data Stewardship Interest Group to send the latest version of the Maturity Matrix to WGCV for review and fine tuning in order to finalize the White Paper that could be presented as a 2023 deliverable to CEOS.**

## [CEOS Interoperability Framework Initiative](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_15.50_CEOS_Interoperability%20Framework.pptx)

Peter Strobl (EC, JRC), Matthew Steventon (Symbios for ESA) gave a presentation on the CEOS Interoperability Framework Initiative. This framework should help coordinate and streamline CEOS efforts, and support by LSI-VC and SIT decisions. Interoperability has several aspects with different requirements:

* Exchange requires technical and organizational connectivity
* Use requires common understanding and legal agreement

Complex issues are best tackled by breaking them down into simpler and related sub-parts to help treat the issue comprehensively (no gaps) and efficiently (no overlaps). At the same time, be (as much as possible) independent from each other and with clear delineations and interfaces and have the right granularity to distribute the effort equally and assemble enough contributors.

Addressing the full scope of interoperability requires a framework that captures all aspects of the problem. The following (non-exhaustive list) are equally important factors to consider, and each have several component pieces:

* Structural / Data Architecture
* Accessibility / System / Data Presentation
* Formats / Syntactic / Data Language
* Terminology / Semantics / Data Context

Factors and components are distributed across many CEOS groups (e.g., LSI-VC, WGISS, WGCV, etc.), and therefore a robust framework and coordination mechanisms are needed to address all the necessary pieces.

Interoperability requires commonalities, i.e., a certain structure or organization. For EO data that is traditionally the processing levels, so far defined as a generic chain of refinement regarding the radiometry (or more general the 'measurand') and the geometry of the (satellite) observation data.

For the discussion of 'Analysis Readiness' of data, a clearer separation of these two 'dimensions’ (measure and/geometry) of processing yields a chance to obtain a transparent scheme in which also recommendations about best possible paths (processing sequences) are feasible. This would be advantageous for defining 'Analysis Ready Data' standards at different processing levels and for their respective interoperability.

Discussion points:

Standards speak different languages; how to engage in this area and align with standards organizations (OGC). How to organize the CEOS presence in the ARD standards OGC: One person in OGC that speaks on behalf of CEOS and not specifically the agency.

Elaboration on point clouds, where disparate datasets can be used together, moving away from standard data packaging approaches and toward point clouds. In data cubing, either you standardize the grid system, or stay with point clouds and do the data cubing on demand.

For EO one must consider that spectral hyperspectral and sounding data that do not have the same sampling.

Also consider interoperability for systems, not only for data.

It is better to use the data as they are. Gridding/binning results in a loss of information. But the idea of processing data as a series of gridded snapshots in time is so ingrained in the way people think about the problem.

Some sensor spaces are difficult to work with. The focal plane has many different spaces.

Except for a few experts, there is no interest in Level 0 and Level 1 data. Every measurement must have an expression of area and time coverage.

## [CEOS Common Online Dictionary: Progress with Joint Activity between WGISS and WGCV](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_16_20_Common%20Online%20Dictionary.pptx)

Peter Strobl (EC, JRC), Katrin Molch (DLR) gave a presentation on the CEOS Common Online Dictionary.

The team has completed the merge of WGCV and WGISS glossaries and NOAA NESDIS Lexicon on a technical level into a temporary Wiki solution ([http://calvalportal.CEOS.org/ca/t-d\_wiki](http://calvalportal.ceos.org/ca/t-d_wiki)). The next task is to harmonise the concepts across domains and build consistent ontologies and taxonomies, with these considerations in mind:

* Intervention during revision of ISO standard 19156 “Geographic information - Observations, measurements and samples” - not successful.
  + Definition of “Observation” includes model outputs, e.g., “simulation”,” forecast”
  + 1:1 use of ISO definitions in CEOS context questionable
* Strategy proposed for next steps.

There are several glossaries existing in EO/Geosciences; all have gaps and inconsistencies, and none show structure or ontology:

* ISO/TC 211 terminology management group: <https://isotc211.geolexica.org/>
* OGC: <https://www.ogc.org/ogc/glossary/>, <http://www.opengis.net/def/glossary/>
* INSPIRE glossary: <https://inspire.ec.europa.eu/glossary>
* CEOS: <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary_v1.2.pdf>
* NASA: <https://earthobservatory.nasa.gov/glossary>

How can CEOS do better?

* Consolidate consensus on the necessity to act (within CEOS and beyond)
* Solicit support and build a coalition of those willing
* Outline scope, principles, and purpose of the new dictionary
* Issue a clear mandate
* Agree on common rules and platforms
* Set realistic short-term (half yearly) and mid-term (1-2 yearly) goals
* Ensure commitment and resources
* Actively organize the transition (e.g., discourage use of ‘one-off linear’ glossaries, encourage collaboration on common glossary)

Proposed terminology good practice:

* A stable and unique entry point (DOI)
* Each term needs to be individually addressable
* Each term needs to be uniquely defined, including unified spelling
* Each definition needs to be complete, i.e., building only on everyday language or other defined terms
* All terms used in a definition which are defined themselves need to be marked and linked (parent-relations)
* All related terms where a term is (re-)used need to be listed and linked (sibling and child relations)
* Circular relations (child becomes parent) should be avoided
* Categorizations need to be unambiguous (no overlaps) and wherever possible complete (no ‘grey zones’)

To develop such a ‘collection of concepts,’ the work would probably take a decade, involving a whole meta-level of discussion, research, and expertise. Taking advantage of work done before by experts in terminology and ontologies, the team would need to approach discussion with initial thinking in place, start with a small number of ‘base terms’ to set the stage, and expand with a few ‘high impact terms’ to show how structure works.

A hierarchy of terms would also have to be built.

Way forward:

* Harmonised terminology is a pre-requisite and a major pillar for any interoperability effort
* EO is a particularly challenging domain due to its multi-disciplinarity
* Current effort cannot be more than a pilot and demonstrator
* Terminology is not stand-alone and needs to be linked with other interoperability efforts
* Ideally as part of a comprehensive Interoperability Framework

Discussion points:

Convergence in these concepts must be achieved in the long term in order to achieve victory.

At the CEOS level they are calling for terminology so this is an opportunity to influence the communities.

This is a critical task for CEOS and for interoperability.

This will not go forward without acceptance.

There was some activity on this at CEOS a couple of years ago. Double sided: if you do not get people involved you will not get adoption. But you cannot have too many people otherwise it is not manageable. What is crucial is for people to be able to read, use, understand, and to provide feedback.

## Greetings from the Senior Chief Officer of Earth Observation Missions

Takeshi Hirabayashi (JAXA), Senior Chief Officer of Earth Observation Missions, Satellite Applications and Operations Center (SAOC), welcomed the participants to Japan and thanked them for their active and productive discussion. He noted that one of the weak points of satellite observations is the limited spatial coverage, so interoperability is indispensable. He appreciated the work of WGCV, including maintaining the cal/val portal and working on ARD, and the work of WGISS in the areas of stewardship and accessibility of data.

# Agency and Liaison Reports



## [Indian Space Research Organization (ISRO)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_11.00_ISRO%20Agency%20Report.pdf)

Nitant Dube (ISRO) gave an agency presentation for the Indian Space Research Organization (ISRO). He highlighted the following points:

EOS-6 (Oceansat-3), INSAT-3DS and NISAR for forthcoming Earth Observation Satellites.

Bhoonidhi-EO Data Hub is now enriched with following data. EOS-04, Re-processed Ortho-rectified products from IRS-1A/1B/1C/1D, Sentinel 1 and 2, NOVASAR, Soumi NPP, Landsat-8, and Landsat-9 (over India).

Bhoonidhi-Vista is a new visualization tool, which helps in visualization of near-real time data.

Bhoonidhi-Upagrah is a real time Satellite Tracker for EO Missions with the ability to track the live location of various in-orbit satellites.

Bhuvan (Portal for Geospatial data and Services) now includes: Evapotranspiration Product available as part of National Hydrology Project, Urban body Information System, Monitoring for recent Floods and Forest Fire data between 2013-2021 (for generation of Forest Fire Zonation Maps).

VEDAS: This web portal is strengthened to enhance web-based geospatial data analysis. New capabilities include Geospatial Query on web, Geospatial calculator, long term trend analysis, RGB composites and site selection tool for Solar power plants. A customized Processing Platform for International Charter on Space and Major Disasters is developed and operationalized.

MOSDAC: Alert and Forewarning System for heavy rain, cloud burst and cyclone is released for Indian Region and South East Asian Countries. The Safe Beach application provides Rip current and Ocean state forecast for 175 beaches of India.

The National Information System for Climate and Environment Studies (NICES) purpose is to Generate and Disseminate Terrestrial, Ocean and Atmospheric Products. Recent additions include ECVs for Lightning, LULC, Water Bodies Fraction and Snow Cover Fraction.

## [Japan Aerospace Exploration Agency (JAXA)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_11.10_AgencyReport_JAXA.pdf)

Yousuke Ikehata (JAXA) gave an agency report on the Japan Aerospace Exploration Agency (JAXA). He described the JAXA portals and data provision to partner portals and reported that the catalog linkage to IDN and FedEO is underway. JAXA is also beginning to assign a DOI to EO datasets.

Yousuke described the G-Portal, JAXA’s main data dissemination system and displayed an image of the architecture of the ground system for EO missions. He also listed JAXA’s policy on intellectual properties, detailing availability, and constraints.

Yousuke gave an overview on ALOS-4, the successor of ALOS-2, and listed the international agency partners for science and applications, with additional details on the JAXA-NASA-ESA cooperation in response to Covid-19.

JAXA also cooperates

for Development of Global Biomass Map

with Google Earth Engine

The Japanese national satellite data platform is “Tellus,” a development and maintenance of data platform aimed for enhancement of satellite data utilization for business purposes launched by the Ministry of Economy, Trade, and Industry of Japan (METI).

JAXA is promoting utilization of satellite data in economics and social science research in universities.

Yousuke concluded with challenges and thoughts:

JAXA is gradually addressing to open-source science and has started with modification or improvement of data dissemination system. JAXA has also begun cooperation with various partners including service providers to promote open science.

Conditions of intellectual properties, particularly software and tools, vary each satellite mission and many software and tools are not intended to be open to the public. Cost of processing of huge volume satellite data, i.e., ALOS-2, is also a challenge to have them available with open and free condition.

Partnership is indispensable for enhancement of further scientific researches and downstream applications for next generation. – JAXA’s IT resources and capabilities are limited to address to integration of big data and AI. In order to have satellite observation incorporated into the society, promoting applications for economics and social sciences is indispensable.

JAXA is interested in NASA’s open-source science initiative as an opportunity to promote scientific researches and applications for next generation.

Richard asked for elaboration on the G-Portal. Yousuke noted that it is for the general user, for example for analyzing typhoons and other disasters.

Valerie reminded that restricted data is not harvested to IDN.

## [United States Geological Survey (USGS)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/5-Wednesday/2022.10.05_11.20_USGS%20Agency%20Report.pdf)

Tom Sohre (USGS) gave a presentation on the United States Geological Survey (USGS). He began with the agency’s organizational structure, and the USGS mission areas.

Tom reported that Landsat 9 was launched 27 September, 2021. He discussed the National Space Council Meeting last December that released the Space Priorities Framework synchronizing civil, commercial, and national security space communities.

A partnership exists between USGS and NASA to ensure sustained access to high-quality, global, land imaging measurements compatible with the existing 50-year Landsat record for research and operational users.

Tom described current Landsat operations status, and Landsat Collection 2. He also described the Landsat Collection 3 preparations.

Tom noted that USGS just celebrated 50 years of Landsat success, and announced Pecora-22, the flagship land-imaging satellite applications conference (a longstanding USGS-NASA partnership).

Mirko asked about the longevity of Landsat-7. Tom replied that it is still collecting data, and there is no date set to stop collecting science data but probably it will be in the next year. USGS is trying to utilize it if it is relevant.

## [German Aerospace Center (DLR)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_14.10_WGISS-DLR-AgencyReport.pdf)

Katrin Molch (DLR) gave a report on the German Aerospace Center (DLR).

DLR is operating three current EO missions:

* TerraSAR-X and TanDEM-X, − Polarimetric, interferometric X-band radar Earth observation − Launched 2007 / 2010 − Public-private partnership DLR / Airbus-DS − Global high-resolution DEM completed in 2016
* EnMAP − Imaging spectrometry − Launched 04/2022 − 242 bands (VNIR, SWIR); 30 m x 30 m spatial resolution − 30 km swath width; max. 5000 km swath length acquired per day − Data dissemination start planned for 11/2022

Four missions are in the pipeline:

* MERLIN − Methane Remote Sensing LIDAR Mission − German-French collaboration − Planned launch 2027
* CO2Image − High-resolution mission complementing Copernicus CO2M − Planned launch 2026
* HRWS − High-resolution Wide Swath − Polarimetric, interferometric X-band radar Earth observation − Launch date is to be determined.
* Tandem-L − Polarimetric, interferometric L-band radar Earth observation − Tomographic measurements of 3-d structure (forest, ice sheets) − Launch date is to be determined.

Katrin discussed Big Data Management in Earth Observation. This involves managing large-volume EO data along the value chain, data modeling, ARD generation, software development, and system and IT engineering. Additional tasks include algorithm development and generation of regional and global geoinformation products, data provision according to international standards, state-of-the-art infrastructures and data services serving EO missions and research projects, and coupling of long-term archive and analysis platforms.

DLR is Developing and operating a range of state-of-the-art infrastructures and services covering the entire EO value adding chain:

* German Satellite Data Archive (D-SDA) Long-term preservation and curation, ensuring EO data accessibility and usability for future users
* EOC-Geoservice Spatial data infrastructure for interoperable access to spatial information products based on international standards
* HPDA Terrabyte High-performance earth observation analytics platform for scientific use in collaboration with the Leibniz Supercomputing Centre, Munich
* Environmental and Crisis Information Systems (UKIS) Set of building blocks for developing customized information and decision support systems
* Copernicus Data and Exploitation Platform CODE-DE 2 − Copernicus Collaborative Ground Segment − EO data and services focused on national public authorities − Funded by the German Federal Ministry for Digital and Transport
* Scientific EO Analytics Platform EO-Lab − EO data and services focused on national scientific users − Funded by the German Federal Ministry for Economic Affairs and Climate Action

DLR also has numerous technical contributions to the Copernicus program. A few examples are:

* Sentinel Core Ground Segment
* Copernicus Collaborative Ground Segment
* Mission Performance (MPC), Instrument and Algorithm Validation, Quality Working Groups (QWG)
* Mission Advisory Groups
* Copernicus Services
* Curated archive of Sentinel-1, Sentinel-2, Sentinel-3 OLCI, Sentinel-5P user level data

Makoto commented that federation of HPC and archival has been discussed and asked for her feedback. Katrin replied that federation is very important to DLR, taking care that proprietary solutions are not installed, and all is interoperable. They have several use cases for hyperspectral satellites.

## [National Aeronautics and Space Administration (NASA)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_14.20_NASA%20Agency%20Report.pdf)

Andrew (Andy) Mitchell (NASA) gave a report on activities at the National Aeronautics and Space Administration (NASA), focusing on open-source science. He highlighted the following points:

* NASA recently established a Chief Science Data Office which is actively implementing many of NASA’s Open-Source Science initiatives.
* NASA Science Divisions are moving towards expansion of cloud activities.
* Multiple Divisions are engaging in new AI/ML activities.
* Divisions are facilitating improved information and knowledge discovery.
* Feedback is received from the community on the implementation of the strategy for Open Science.
* Tangible activities are identified for collaboration and development between NASA science data repositories.
* Networks are built to support the Open Science vision.

The NASA Science Mission Directorate strategy is to enable transformational open science through continuous evolution of science data and computing systems. The goals are to develop and implement capabilities to enable Open Science with a continuous evolution of data and computing systems, harnessing the community and strategic partnerships for innovation.

Andy described the Scientific Information Policy and described several ongoing projects:

* The science discovery engine.
* The NASA Astrophysics Data System is a NASA-funded project which provides discovery services for scholarly literature in astronomy and physics.
* Transform to Open Science is a 5-year NASA SMD initiative to foster adoption of Open Science practices across the scientific community.
* Context for Data and Computing Architecture Study is conducting two activities to develop cyberinfrastructure to support the Strategy for Data Management and Computing and the science data policy.

Ken requested more information on the Science Discovery Engine software. Andy replied that they will be aggregating the software, using a commercial off-the-shelf software package that will be connected to the CMR. It will be the heart of how data is fed to data.gov.

## Agencia Espacial Mexicana

Adrian Guzman (Agencia Espacial Mexicana) introduced himself as a representative from international relations at the agency. The agency is working on an initiative for Latin American space agencies and two regional data centers. As the WGISS principal, he has a technical background and is a heavy user of opensource and having a good physical infrastructure . He is interested in collaboration with WGISS.

# WGISS Plenary, Part II



## [Future Meetings](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_16.30_Future%20Meetings.pptx)

Tom Sohre (USGS) announced the planned future WGISS meetings:

* WGISS-55: Proposed dates 18-20 April 2023, hosted by Comisión Nacional de Actividades Espaciales (CONAE), in Buenos Aires, Argentina (hybrid format).
* WGISS-56: Proposed dates 3-5 October 2023, hosted by Centre National d’Etudes Spatiales (CNES), Paris, France (hybrid format).

Tom provided detailed information about travel and venue for WGISS-55.

**Action WGISS-54-16: Michelle Piepgrass to add the information on WGISS-55 and WGISS-56 to the CEOS calendar and the WGISS website. Due October 31, 2022.**

## [WGISS Vice Chair Solicitation](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_16.40_vice%20Chair%20Solicitation.pptx)

Makoto Natsuisaka (JAXA) formally solicited a vice-chair for the period 2023-2025. The current WGISS Chair Team consists of:

　 Chair: Makoto Natsuisaka (JAXA)

　 Vice-chair: Tom Sohre (USGS)

Secretary: Michelle Piepgrass (JAXA)

The next chair team will begin their activities right after the 2023 CEOS Plenary. The new vice-chair nomination should be done in advance of the 2023 CEOS Plenary.

Chair: Tom Sohre (USGS)

Vice-chair: TBD

The process to follow is:

Solicitation: WGISS-54 (October 2022)

WGISS nomination: WGISS-55 (April 2023)

CEOS nomination: 2023 CEOS Plenary (October 2023)

Vice-chair requires a 4-year commitment (beginning after the October/November 2023 CEOS Plenary)

* Initial 2 years as Vice-chair
* Next 2 years as Chair (beginning after the October/November 2025 CEOS Plenary)

Chair and Vice chair meeting attendance expectations:

* CEOS Plenary (once a year, October/November)
* CEOS SIT (once a year, March/April)
* CEOS SIT Technical Workshop (once a year, September/October)
* CEOS SEC (every month, remote)
* WGISS Plenary meetings (twice a year)
* WGISS Exec Meetings (every month, remote)

Additional WGISS Chair tasks:

* Lead WGISS activities
* CEOS Work Plan update (once a year)
* CEOS deliverables update (once/twice a year)
* WGISS report at CEOS Plenary
* Maintain Secretary Service

Makoto stated that agency candidacy would be welcome; please contact [natsuisaka.makoto@jaxa.jp](mailto:natsuisaka.makoto@jaxa.jp). The solicitation will be open until the WGISS-55.

Andy suggested that recruiting for WGISS participation across CEOS agencies would be helpful, perhaps targeting the WGISS list of principals. Esther added that activities like Jupyter Notebooks Day would provide additional possibilities.

**Action WGISS-54-17: Michelle Piepgrass to circulate the WGISS-All and WGISS-Internal email lists for review and update by November 30, 2022.**

## [WGISS Summary and Discussion](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/6-Thursday/2022.10.06_16.50_WGISS%20Summary%20and%20Discussion.pptx)

Makoto Natsuisaka (JAXA) summarized the highlights of this meeting as follows:

Plenary:

* Need to update WGISS resources on WGISS website to align with the CEOS work plan: Organization chart, ToR, what can WGISS do for CEOS agencies and Data Providers?
* Solicit new vice-chair (after 2023 CEOS Plenary until 2025 CEOS Plenary) by the end of December 2022. The assignment will be 4-year commitment including 2-year chair term.
* The activities of the CEOS ARD Oversight Group and the CEOS Ocean Coordination Group were introduced.
* CEO and SEO reports were given.
* GEO presentation on GEO work plan, GEO knowledge hub were introduced.
* ISO TC211/OGC liaison report was given. ARD activity has already started.
* WGISS-55 will be held the of 18th – 20th April in CONAE.

Data Preservation and Stewardship:

* Archive Technology and Future Trends were reported.
* WGISS DSIG Best Practices Refreshment were proposed.
* The results of the information exchange with CGMS IV group were reported. Common interests on MM were identified.

Data Discovery and Access:

* IDN and FedEO status were reported.
* OGC Testbed-18 status was reported.
* Service Metadata and Discovery Best Practices document was outlined and proposed.

Data Interoperability and Use:

* EO Interoperability for Services was presented. Problems on data handling between HPC and archive were pointed out.
* EAIL use case of COAST was introduced. The EAIL was used for shoreline mapping and continuity of EAIL was requested.
* Status of development of EAIL was reported.
* STAC related issues were presented.
* OGC STAC CMR implementation was proposed and volunteer from CWIC provider was solicited.
* STAC Best Practice is under development.

Technology Exploration:

* Jupyter Notebook initiative is ongoing. Jupyter Notebooks Day will be held on 21st Oct. as a collaboration with SEO and WGCapD. The expectation for Jupyter Notebooks initiative from WGCapD was expressed.
* Informative presentations on AI/ML were given.
* NASA add keywords related to AI/ML to GCMD keywords.
* NASA adapted knowledge graph to CMR.
* NOAA has already put effort into AI/ML.
* Study of federation between HPC and Archive were proposed. White paper would be a target.

WGISS/WGCV Joint Symposium:

* Sixth joint symposium with WGISS-WGCV was successfully held.
* ESA EO Portal, CEOS MIM Database: Status, Synergies and Improvements, CEOS Cal/Val Portal and Status of ARD Activities were reported.
* Data Quality Assessment and Indicators, and DMSMM Maturity Matrix were reported. The information exchange between WGISS and WGCV will be done toward issue of the Best Practice planned in the next year.
* New activities “CEOS Interoperability Framework Initiative” and “CEOS Common Online Dictionary” were proposed as way forwards of CEOS ARD.

## [Review of WGISS Actions](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-54/WGISS-54%20Actions.xlsx)

Michelle Piepgrass (JAXA) and the participants discussed the actions from this meeting:

Action WGISS-54-01: Makoto Natsuisaka and WGISS Exec to begin drafting a list of CEOS Workplan deliverables for the period 2023-2025. Due January/February 2023.

Action WGISS-54-02: WGISS Exec to request from the SEO a copy of the charter that describes the goals and objectives of cloud prototyping efforts. If a charter does not exist, create a follow-up action to work with SEO to charter the efforts so that CEOS has defined goals and objectives.

Action WGISS-54-03: Iolanda Maggio and Mirko Albani to participate in the discussion with GEO on 24th of October regarding the self-assessment tool in the GEO Knowledge Hub, to share experiences based on the WGISS Data Management and Stewardship Maturity Matrix which is based on the GEO Data Management Principles.

Action WGISS-54-04: Data Stewardship Interest Group to issue a plan for data stewardship best practices refreshment with priorities and schedule. Due by October 2022.

Action WGISS-54-05: WGISS Exec to review prior browse data best practices and guidelines, and consider updating to reflect newer cloud capabilities.

Action WGISS-54-06: Mirko Albani to organize a session at WGISS-55 on AVHRR data recovery including dedicated presentations on European and North America LAC Archives and progress on other worldwide archives. Due by WGISS-55.

Action WGISS-54-07: Mirko Albani to contact CGMS rapporteur (Simon Elliott, EUMETSAT) to understand if/how a join session with WGISS could be organized. (Note: CGMS WG-IV will have dedicated meetings in January 2023 and March 2023; CGMS chair is Kotaro Bessho, kbessho@met.kishou.go.jp). Due by October 2022.

Action WGISS-54-08: Data Stewardship Interest Group to evaluate potential extension/application of existing WGISS data preservation guidelines to ground-based (in-situ) data. Due by WGISS-55.

Action WGISS-54-09: WGISS Exec to define a use case that clarifies the challenges of authenticating data replicas in the cloud, including related goals, options, and technologies. Due by WGISS-55.

Action WGISS-54-10: Mirko Albani to organize a dedicated presentation at WGISS-55 on Copernicus Sentinels Ground segment architecture and authenticity and provenance approaches, inviting outside participants. Due by WGISS-55.

Action WGISS-54-11: WGISS Exec to study an approach to define and obtain useful and valid metrics of agency data use and access in the Cloud that quantify measures of success. Due by WGISS-55.

Action WGISS-54-12: Makoto Natsuisaka to share CEOS Interoperability Framework to WGISS, and call for participants.

Action WGISS-54-13: Technology Exploration Interest Group to add the production and publication of the AI/ML White Paper to the CEOS Work Plan. Due by WGISS-57 (April 2024).

Action WGISS-54-14: Technology Exploration Interest Group to develop in-depth presentations at WGISS-55 on federation, including a call for proposals for federation. Due WGISS-55.

Action WGISS-54-15: Data Stewardship Interest Group to send the latest version of the Maturity Matrix to WGCV for review and fine tuning in order to finalize the White Paper that could be presented as a 2023 deliverable to CEOS.

Action WGISS-54-16: Michelle Piepgrass to add the information on WGISS-55 and WGISS-56 to the CEOS calendar and the WGISS website. Due October 31, 2022.

Action WGISS-54-17: Michelle Piepgrass to circulate the WGISS-All and WGISS-Internal email lists for review and update by November 30, 2022.

## Concluding Discussion and Remarks

Makoto Natsuisaka (JAXA) thanked the attendees for their participation and fruitful presentations and discussions, and adjourned the meeting.

# Glossary of Acronyms

API Application Programming Interface

ARD Analysis Ready Data

AWS Amazon Web Services

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

COTS Commercial Off-the-Shelf

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

DC data cube

DIF Directory Interchange Format

DOI Digital Object Identifier

ECV Essential Climate Variable

EO Earth Observation

ESIP Federation of Earth Science Information Partners

GCI GEOSS Common Infrastructure

GCMD Global Change Master Directory

GEO Group on Earth Observations

GEO-GLAM Global Agricultural Monitoring

GEOSS Global Earth Observation System of Systems

GFOI Global Forest Observations Initiative

GHG Greenhouse Gas

GIS Geospatial Information System

GPM Global Precipitation Mission

GPU Graphics Processing Unit

GSDI Global Spatial Data Infrastructure

GUI Graphical User Interface

HPC High Performance Computing

ICT Information and Communication Technology

IDN International Directory Network

ISO International Standards Organization

LSI Land Surface Imaging

LTO Linear Tape-Open

MM Maturity Matrix

MOU Memorandum of Understanding

NRT Near real-time

NWIP New Work Item Proposal

OGC Open Geospatial Consortium

PI Persistent Identifier

POC Point of Contac

RS Remote Sensing

SEO Systems Engineering Office

SDCG Space Data Coordination Group

SIT Strategic Implementation Team

SLT System Level Team

SWG Standards Working Group.

TEP Thematic Exploitation Platform

ToR Terms of Reference

UML Unified Modelling Language

UMM Unified Metadata Model

VC Virtual Constellation

WCS Web Coverage Service

WG Working Group

WGCV Working Group on Calibration and Validation

WGCapD Working Group on Capacity Building and Data Democracy

WGClimate Working Group on Climate

WGDisasters Working Group on Disasters