MINUTES

OF THE

55th MEETING

OF THE

CEOS WORKING GROUP ON   
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Córdoba, Argentina

18-20 April, 2023

Hosted by

Comisión Nacional de Actividades Espaciales (CONAE)

Table of Contents

[1 WGISS Plenary Session, Part I 7](#_Toc135300199)

[1.1 Introduction 7](#_Toc135300200)

[1.2 Welcome from CONAE and Logistics 7](#_Toc135300201)

[1.3 WGISS Chair Report 7](#_Toc135300202)

[1.4 CEOS Executive Officer Report 8](#_Toc135300203)

[1.5 Discussion of CEOS Work Plan Deliverables for WGISS 9](#_Toc135300204)

[2 CEOS Interoperability Framework 10](#_Toc135300205)

[2.1 Session Objectives 10](#_Toc135300207)

[2.2 Introduction to the CEOS Interoperability Framework 10](#_Toc135300208)

[2.3 Professional Open-Source Framework for Earth System Digital Twins and Applications 11](#_Toc135300209)

[2.4 Cloud Use Activities and Strategy 13](#_Toc135300210)

[2.5 Discussion on the Vision, Outcomes, Actions, Next Steps 13](#_Toc135300211)

[3 WGISS-WGDisasters Joint Symposium 15](#_Toc135300212)

[3.1 Introduction to CONAE 15](#_Toc135300214)

[3.2 Overview of WGISS Activities 15](#_Toc135300215)

[3.3 Overview of WGDisasters Activities 15](#_Toc135300216)

[3.4 AI/ML 16](#_Toc135300217)

[3.4.1 Presentation of WGISS AI/ML White Paper 16](#_Toc135300218)

[3.4.2 Presentation of WG Disasters Planned Use Case on ML and the Wildfire Sector 16](#_Toc135300219)

[3.5 Service Discovery Best Practice Activity 17](#_Toc135300220)

[3.6 Use of EAIL in Flood Pilot and Other Possible Synergies 17](#_Toc135300221)

[3.7 Outcomes, Actions, Next Steps 17](#_Toc135300222)

[4 Data DISCOVERY and ACCESS 18](#_Toc135300223)

[4.1 Session Objectives 18](#_Toc135300225)

[4.2 Data Access Authorisation - Toward Federated Access 18](#_Toc135300226)

[4.2.1 EGI (OpenEO) – Federated Authorisation 18](#_Toc135300227)

[4.2.2 NASA - Multifactor Authorisation 18](#_Toc135300228)

[4.3 GEO-ZARR Software Working Group 18](#_Toc135300229)

[4.4 ARD Standardization in ISO TC211 and OGC 19](#_Toc135300230)

[4.5 IDN and CWIC Update 19](#_Toc135300231)

[4.6 FedEO Update 20](#_Toc135300232)

[4.7 Metrics for Cloud-based Data Use 20](#_Toc135300233)

[4.8 Service Discovery: OSS Tools Metadata Ingestion Use Cases – FedEO 21](#_Toc135300234)

[4.9 Outcomes, Actions, Next Steps 21](#_Toc135300235)

[5 Data PRESERVATION and STEWARDSHIP 22](#_Toc135300236)

[5.1 Session Objectives 22](#_Toc135300238)

[5.2 Archive Technologies Evolution White Paper 22](#_Toc135300239)

[5.3 DSIG Best Practices Refreshment 24](#_Toc135300240)

[5.4 Copernicus Sentinels Ground Segment Architecture 25](#_Toc135300241)

[5.5 Session on AVHRR Data Archives 26](#_Toc135300242)

[5.5.1 AVHRR Data Recovery Project 26](#_Toc135300243)

[5.5.2 AVHRR Data: European Coverage 26](#_Toc135300244)

[5.5.3 AVHRR Data: North American Coverage 27](#_Toc135300245)

[5.6 Introduction of DOIs to the JAXA G-Portal Products 27](#_Toc135300246)

[5.7 Data Replicas (Authenticity and Integrity) in the Cloud Use Case 27](#_Toc135300247)

[5.8 Authenticating Data Replicas 28](#_Toc135300248)

[6 TECHNOLOGY EXPLORATION 29](#_Toc135300249)

[6.1 Session Objectives 29](#_Toc135300251)

[6.2 Status of Jupyter Notebooks Best Practice 29](#_Toc135300252)

[6.3 Artificial Intelligence/Machine Learning (AI/ML) 30](#_Toc135300253)

[6.3.1 Proposal for AI/ML Survey in WGISS 30](#_Toc135300254)

[6.3.2 NOAA Steps to the Geoverse 30](#_Toc135300255)

[6.3.3 Plymouth Marine Lab Work 31](#_Toc135300256)

[6.3.4 MLHub and AI/ML in CMR 31](#_Toc135300257)

[6.4 Federation 32](#_Toc135300258)

[6.4.1 STAC Best Practice Activity 32](#_Toc135300259)

[6.5 Outcomes, Actions, Next Steps 33](#_Toc135300260)

[7 Agency and Liaison Reports 34](#_Toc135300261)

[7.1 ASI (Italian Space Agency) 34](#_Toc135300263)

[7.2 European Space Agency (ESA) 34](#_Toc135300264)

[7.3 Indian Space Research Organization (ISRO) 34](#_Toc135300265)

[7.4 NOAA 35](#_Toc135300266)

[7.5 Russian Federal Space Agency (ROSCOSMOS) 35](#_Toc135300267)

[7.6 United Kingdom Space Agency (UKSA) 35](#_Toc135300268)

[7.7 USGS Emergency Operations Activities 36](#_Toc135300269)

[8 WGISS Plenary, Part II 38](#_Toc135300270)

[8.1 WGISS Strategic Discussion of the CEOS Interoperability Framework 38](#_Toc135300272)

[8.2 Future Meetings 40](#_Toc135300273)

[8.3 WGISS Vice Chair Solicitation, and Organizational Matters 40](#_Toc135300274)

[8.4 Review of WGISS Actions 40](#_Toc135300275)

[8.5 Concluding Discussion 41](#_Toc135300276)

[8.6 Meeting Conclusion 42](#_Toc135300277)

[9 Glossary of Acronyms 43](#_Toc135300278)

Agencia Espacial Mexicana Adrian Guzman\*

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

CEOS Marie-Claire Greening (CEOS Executive Officer)

CNES Richard Moreno\*, Pierre-Marie Brunet\*, Than Huy Nguyen\*

CONAE Homero Lozza

DLR Katrin Molch\*

ESA Mirko Albani, Roberto Alacevich\*, Barbara Angelucci\*, Damiano Guerrucci\*, Daniele Iozzino (Rhea)\*, Brian Keary\*, Iolanda Maggio (Rhea), Filippo Marchesi (Solenix)\*, Christophe Noel (Spacebel)\*, Sabrina Pinori (Serco)\*, Paulo Sacramento\*, Sean Walsh\*

GEO Bernd Eversmann\*

ISO TC 211 Liping Di\*

ISRO Nitant Dube\*, Sai Kalpana Tanguturu\*

LSI-VC Matt Steventon\*

JAXA Makoto Natsuisaka (WGISS Chair), Yousuke Ikehata, Michelle Piepgrass, (WGISS

secretary)

NASA Andrew (Andy) Mitchell\*, Dana Ostrenga\*, David Borges (SEO)\*, Diane Davies\*, Valerie Dixon\*, Brian Killough\*,Thomas Huang, Dawn Lowe\*, Michael Morahan, Doug Newman, Patrick Quinn\*, Hampapuram Ramapriyan (Rama)\*, Kenton Ross\*, Min Wong

NOAA Kenneth (Ken) Casey\*, Rich Baldwin\*, Ryan Berkheimer\*

NRCan Alan Cantin\*, Joshua Johnston\*

NRSCC Chuang Liu\*

UKSA Esther Conway, David Moffat (PML)\*

UniBern Stefan Wunderlee\*

USGS Thomas (Tom) Sohre (WGISS Vice-chair)

WGDisasters Laura Frulla, Andrew Eddy, Andrey Kuklin,

**GROUP PHOTO**

****

# WGISS Plenary Session, Part I

## Introduction

Makoto Natsuisaka (WGISS-Chair, JAXA) opened the WGISS-55 meeting, thanking the participants for their attendance, and CONAE for hosting the meeting. He noted that this is a working meeting, and the agenda contains many opportunities for discussion. He asked the in-person and remote participants to introduce themselves. Makoto shared the following highlights of the meeting:

* Share information on CEOS activities
* CEOS Work Plan 2023-2025, CEO Report
* Joint session with WGDisasters to foster mutual understanding, identify seeds of collaboration (AI/ML, Service Discovery, EAIL)
* CEOS Interoperability Framework: Discuss WGISS vision, contributions, and contributors, and how this activity impacts the WGISS structure
* CEOS New Space Task Team (NSTT): WGISS contributions
* Interest group sessions; status reports and discussions on the way forward
* Member reports; foster mutual understanding among WGISS and WGDisasters members, CONAE, and ISO/TC211
* Discussion with SEO on Cloud use and SEO strategy and desire for collaboration with WGISS
* WGISS organizational matters: Vice-chair solicitation, re-formation of Interoperability Data and Use Interest Group, WGISS vision and feedback to CEOS Workplan and WGISS Terms of Reference.

## Welcome from CONAE and Logistics

Homero Lozza (CONAE) and Laura Frulla (CONAE) welcomed everyone to Argentina and to CONAE and described the logistics for transportation, meals, breaks, and organized events.

## [WGISS Chair Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_09.30_WGISS%20Chair%20Report.pptx)

Makoto Natsuisaka (JAXA) reported on the following:

* 2022 CEOS Plenary (Nov. 29 – Dec. 01, 2022, Biarritz, France)
* CEOS Chair (CNES) decision to develop the CEOS Interoperability Framework
* CEOS Work Plan 2023-2025 submitted (CEO report)
* CEOS ARD Oversight Group (ISO TC211 report)
* CEOS Ocean Coordination Group: This is a task force formally led by the SIT Chair and currently done by the Coast AHT Chair. The objective is to coordinate CEOS ocean-related activities and clarify CEOS’ contribution to the UN Ocean Decade. A six-month extension for further discussions was approved at SIT-38. WGISS Chair was invited to the activities regarding future use of EAIL, but the operation of EAIL was moved to the joint activity by SEO/CSIRO and discussion points are thematic, and accordingly the WGISS contributions are considered as completed.
* CEOS Engagement with Standard Organizations: The issue was raised at SIT-TW 2022; in discussions around standards for Analysis-Ready Data, it became clear that CEOS was perhaps uncoordinated in its approach to engagement with standards bodies. For ARD, this coordination has since been resolved via the ongoing effort of the CEOS-ARD Oversight Group. Question remains whether CEOS generally needs more formal coordination around engagement with standards bodies, e.g., ISO/OGC/IEEE. Consultation with the CEOS community at 2022 SIT Technical Workshop side meeting – agreed that more coordination on the engagement of CEOS with standards organizations would be beneficial. CEOS Plenary discussion and results: SIT Chair to organise a dedicated session on the topic of CEOS engagement with standards organisations at SIT-38 and invite CEOS entities to present their experience and status, with the aim of creating a snapshot inventory.

WGISS provide input to SIT-38 based on the discussions in the WGISS Exec meetings in 2022 as follows: WGISS solicited a volunteer for the activity at the WGISS EXEC meetings. WGISS does not have experts on standardizations except for Liping Di, who is a liaison with OGC/ISO-TC211, nor does it have resources. The direct commitment to the activity is difficult, but WGISS will support the activity through regular members of OGC/ISO-TC211 in individual agencies.

* CEOS New Space Task Team was established for a period of one year at 2022 CEOS Plenary. Objectives:

Objective 1: Explore collaboration opportunities in New Space that bring mutual benefit to all parties, in delivering public benefit, including the identification of concrete actions. Define and propose actions within the CEOS framework that enhance the outcomes of CEOS entities (WGs, VCs, AHTs) working with New Space companies. Foster partnership at national level between individual CEOS members and their national “New Space” industry. Focus on the whole value chain – upstream and downstream sectors

Objective 2: Continue sharing country-level experience among CEOS. Commence with an assessment of the areas and issues that are common among CEOS agencies and that are predicted to impact public EO programs in future. Develop a series of objectives and activities that will guide CEOS in fostering the combined use of public and private EO New Space capabilities, both upstream and downstream. SIT-38: Continue agency sharing of experience from SIT-37 and TW, and accelerate the definition of actions. 2023 Plenary: issue a white paper with findings and recommendations. The white paper will highlight which topics are most productive for CEOS to pursue and which may be out of scope or problematic.

* CEOS Interoperability Framework (CEOS Interoperability Framework session).
* SIT-38 (Mar. 28-30, 2023, ESA/ESRIN, Frascati, Italy)

## [CEOS Executive Officer Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_09.45_CEOReport.pptx)

Marie-Claire Greening (CEO) gave the CEOS Executive Officer report. She thanked CONAE for hosting the meeting, and WGISS for inviting her. She began with a reminder of the CEOS mission and objectives, and its long-term priorities. CEOS is fortunate to have good governance to support the work.

The GISTDA CEOS Chair key priorities are:

* Supporting CEOS Preparations and Inputs to the Global Stocktake of the UNFCCC Paris Agreement
* Supporting Exploration of New Geometries for Space Agencies and CEOS with New Space

Marie-Claire reminded of the WGISS mission (To facilitate (CEOS) data and information management, and to support the provision of services for both data providers and data users) and objectives:

* To enable Earth observation data and information to be more accessible and usable to both data providers and data users world-wide through international coordination.
* To foster easier exchange of Earth observation and related data and information to meet the requirements of users and data providers.
* To foster the development of best practices and encourage the development of interoperable services that exploit space-borne Earth observation data.
* To enhance the complementarity, interoperability and standardization of Earth observation data and information management and services with other types of geospatial data such as in-situ data.

The four key governing documents of CEOS are:

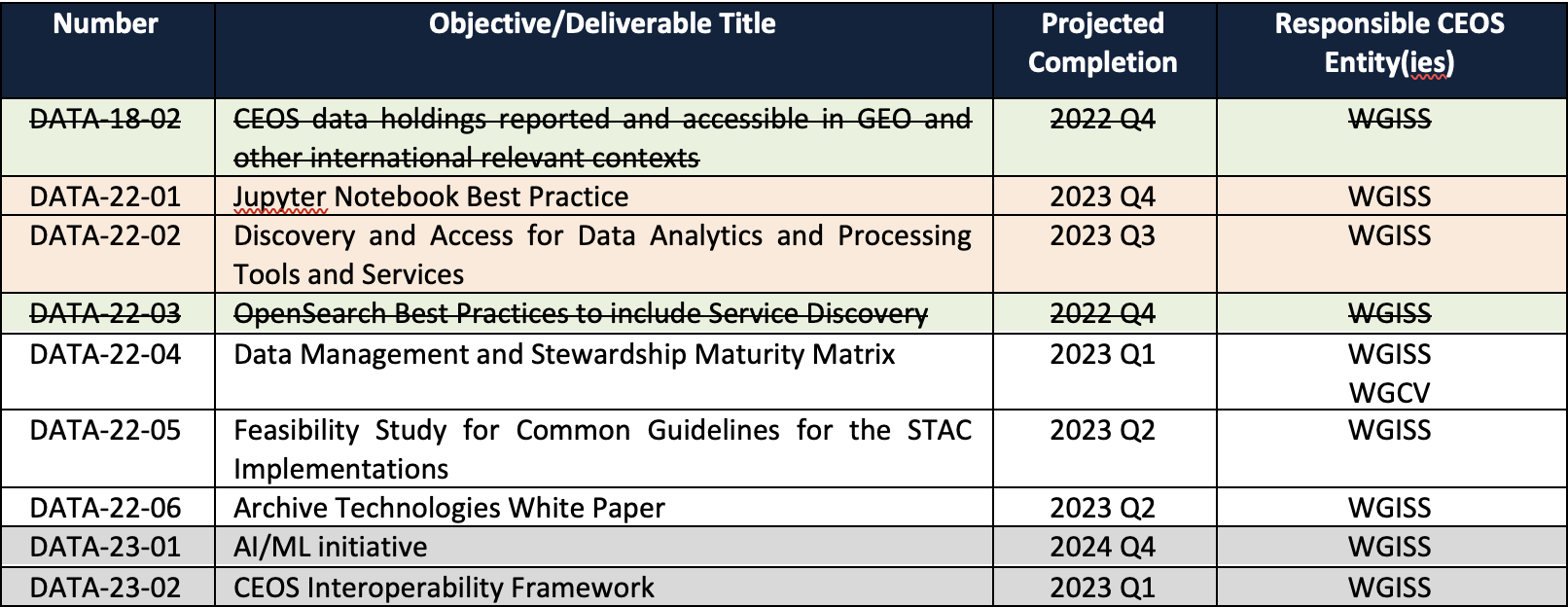
1. CEOS Terms of Reference
2. CEOS Strategic Guidance
3. CEOS Governance and Processes
4. CEOS Work Plan which includes the CEOS deliverables.

Marie-Claire displayed the outline of the 2023-2025 Work Plan. Chapter 9 is dedicated to WGISS: Data Discovery, Access, Preservation, Usability and Exploitation approaches, systems, tools, and technologies.

Marie-Claire confirmed that CEOS regularly coordinates with GEO, and several CEOS members are representatives at GEO. There is also a constant engagement with WMO, and outreach to UN bodies.

## [Discussion of CEOS Work Plan Deliverables for WGISS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_10.00_Discussion%20of%20CEOS%20Work%20Plan%20Deliverables%20for%20WGISS.pptx)

The CEOS Work Plan, chapter 3.9, states: Data Discovery, Access, Preservation, Usability and Exploitation: approaches, systems, tools, and technologies. Through the CEOS Working Group on Information Systems and Services (WGISS), CEOS Agencies will continue to foster the enhancement of the WGISS Connected Data Assets Infrastructure to support not only the CEOS entities, but also the external entities like GEO, UN, WMO, etc., by enabling discovery and access capabilities to mature data services provided by CEOS Agencies.

The WGISS deliverables are listed below. The color code indicates if the deliverable is new, closed, due now, due later:

DATA-18-02: This has been an ongoing activity, and it was confirmed at the end of 2022 that CEOS data holdings are accessible in GEO.

DATA-22-01: The Technology Exploration Interest Groups will discuss the plan for this document in this meeting. Webinars are also being considered.

DATA-22-02: This will be discussed in the Data Discovery and Access session.

DATA-22-03: The Service Discovery Best Practice was released in 2022.

DATA-22-04: The matrix has been submitted to WGCV and is in final review.

DATA-22-05: A survey was distributed and results will be presented in this meeting.

DATA-22-06: The team has collected all the materials needed to write the white paper, and will send it to WGISS for review.

DATA-23-01: Work is ongoing and planning will be discussed in this meeting.

DATA-23-02: Work is ongoing and planning will be discussed in this meeting.

Tom requested further discussion on whether WGISS is organized to handle these deliverables well, given the potential for overlap. For example, Jupyter Notebooks overlaps with nearly everything. Cloud and Interoperability do not necessarily fit in the current boxes.

# CEOS Interoperability Framework



## [Session Objectives](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_10.30_CEOS%20Interoperability%20Framework%20Session%20Objectives.pptx)

Makoto Natsuisaka (JAXA) stated that the session objectives are to share the current status and the draft roadmap, and to discuss the focus of the framework, the factors (interoperability elements), WGISS contributions and those of other entities (including SEO regarding cloud issues), and to consider a timeline of 2023-25 for the roadmap.

As background, Makoto explained that the CEOS Interoperability Framework was proposed at the 2022 SIT Technical Workshop by LSI-VC. The Chair decision at the CEOS Plenary 2022 was that “CEOS Principals confirm that coordination of interoperability-related work remains within the WGISS perimeter”. The following two actions resulted:

* CEOS-36-10: WGISS invited to propose an interoperability roadmap at SIT 38. Any other CEOS group/VC which wants to contribute is invited to contact WGISS.
* CEOS-36-11: CEOS Agencies to consider nominating individuals to the ongoing WGCV / WGISS / LSI-VC effort to define a CEOS common dictionary of terms.

WGISS invited contributors from other CEOS entities and organized the framework development team, to consist of:

Lead: Makoto Natsuisaka (WGISS)

Contributors: WGISS: Tom Sohre

WGCV: Philippe Goryl, Paolo Castracane

LSI-VC: Steve Labahn, Peter Strobl, Ake Rosenqvist

CEOS-ARD OG: Ferran Gascon

SEO: Dave Borges

During February and March WGISS, LSI-VC, CEOS ARD OG, WGCV, and SEO met several times to discuss vision, teams, topics, membership, process, components, and topics. A draft roadmap was proposed by Tom Sohre and Steve Labahn(USGS) , followed by discussion of the components and the topics based on the roadmap. Further discussions in each VC/WG were requested. Makoto requested and received an extension of the roadmap delivery at the 2023 SIT-TW.

## [Introduction to the CEOS Interoperability Framework](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_10.40_CEOS%20Interoperability%20Framework%20Intro.pptx)

Tom Sohre (USGS) gave an introduction to the CEOS Interoperability Framework. The framework would comprise multiple “factors” of interoperability, where all the factors work together. The factors (see image below) were suggested and peer reviewed through a team effort. Each factor will have a lead (from a CEOS WG), but contributions will be necessary from multiple groups. WGISS will have overall leadership role for the Framework, in addition to lead role within several factors.

The next objective is to develop the roadmap of activities within each factor, identifying key 2023-2025 activities aligned to work plans by June 2023 and present the overall roadmap by end of September 2023 (at SIT-TW). The future concept is to develop a Maturity Matrix that includes all Factors.

Tom suggested the following activities:

* Further refinement of scope (in/out) for each factor
* Communication/coordination approach
* Finalize for each factor (coordinated by Lead) including the following items:

Factor name

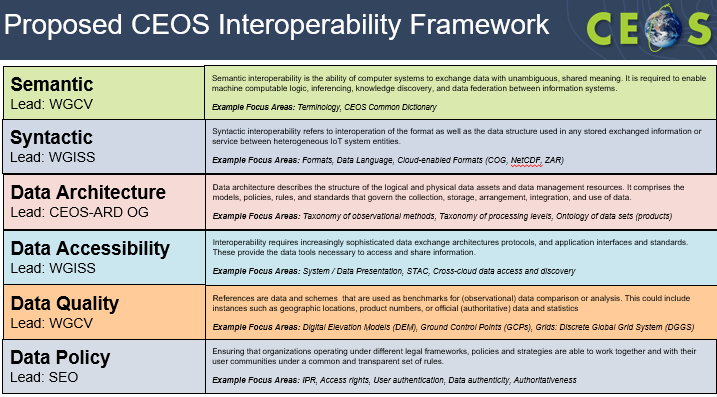
Coordinator (Lead)

Collaborator(s)

Description: 1-2 paragraphs

Scope included

Scope excluded: If relevant for interoperability, and if possible, identify where the item is included



Tom proposed the following discussion topics:

* WGISS to adopt current draft Framework, with concurrence of proposed (6) “factors”, and identify a lead to each.
* WGISS to determine how the Framework will be led: two possible options are WGISS Chair Team or the revived WGISS Interoperability Interest Group. The vision of the lead would be coordination of Roadmap development and status; communications.
* Roadmap development for WGISS-specific roadmap factors, (Syntactic, Data Accessibility), determining which interest groups are responsible, and how to build upon the 2008 Interoperability Handbook.

Yves asked for clarification of the factor “syntactic” – Tom replied that the factor leads will define that, including what is and is not within scope. Ken commented that it is great to build on previous work, given that interoperability encompasses everything WGISS does.

## [Professional Open-Source Framework for Earth System Digital Twins and Applications](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_11.00_Professional%20Open-Source%20Framework%20for%20Earth%20System%20Digital%20Twins%20and%20Applications.pdf)

Thomas Huang (JPL for NASA) gave a presentation on the Professional Open-Source Framework for Earth System Digital Twins and Applications. He began by noting that big data is about being smarter about data, with agility, relevancy, and sustainability.

An Earth System Digital Twin (ESDT) is an interactive and integrated multidomain, multiscale, digital replica of the state and temporal evolution of Earth systems, with a vision of technology to bridge the physical and digital environments. It dynamically integrates:

* Relevant Earth system models and simulations
* Other relevant models (e.g., related to world’s infrastructure)
* Continuous and timely (including near real-time and direct readout) observations (e.g., space, air, ground, over/underwater, Internet of Things (IoT), socioeconomic)
* Long-time records
* Analytics and artificial intelligence tools

An ESDT enables users to run hypothetical scenarios to improve the understanding, prediction and mitigation/response to Earth system processes, natural phenomena, and human activities as well as their many interactions.

Key components of the ESDT:

* Data and Services Assets: Supports Extract, Transform, and Load (ETL) workflow for metadata harvesting, error detection and correction, re-gridding/reprojecting, Analysis Ready Data (ARD) transformation
* New Observation and Analysis: Smarter method to automate onboarding relevant data
* Integrated Multiphysics, Multi-scale, Probabilistic Models: Automates inclusion of the latest measurements and supports scenario-based model execution
* AI and Advanced Analytics: enables dynamic data acquisition, long-term prediction, data classification, process orchestration and management, etc.

For hydrology, flood prediction, and analysis, Thomas presented a federated digital twins solution between the NASA ESTO/AIST’s Integrated Digital Earth Analysis System (IDEAS) (Huang/JPL) and the Space for Climate Observatory (SCO) FloodDAM-DT (Rodriguez-Suquel/CNES). This is built for scenario-based analysis using the latest observation and analysis to drive model predictions, decision support, on-demand data and analysis acquisition, and future instrument scheduling and tasking.

One goal is to minimize storage and computation need for pre-staging different in-situ sensor data, using real-time ML to predict which stream gauges will be most useful for analysis.

The IDEAS-powered flood notebook demonstrates the latest IDEAS API and capabilities.

Thomas also listed the SCO FloodDAM-DT subsystems, and displayed the pipeline and schedule.

For wildland fire, air quality, and health impact, Thomas noted that the goal is to improve usability of science data for air quality analysis and prediction, to leverage advanced assimilation of numerical and AI models to improve decision making, and to develop sustainable technology solution for sustainable science. The air quality data analysis tool is an interactive analysis and visualization of satellite observations, model for atmospheric composition, and IOT and in-situ sensors. The air quality notebook demonstrates the latest IDEAS API for air quality analysis. ML-driven data acquisition uses satellite and in-situ data inputs in predictive models to drive decisions on which in-situ stream gauge readings to ingest.

The open-source science and community collaboration is a partnership with Apache Software Foundation to embrace open-source software, to evolve the technology through community contributions, to host webinars, hands-on cloud analytics workshops and hackathons, and to use Open-Source Science to:

* Share recipes and lessons learned
* Community validation
* Technology demonstrations
* Inclusive and diverse Project Management Committee (PMC)

Thomas presented the following conclusions:

* Building toward Earth System Digital Twin is an opportunity to define and develop reusable, open-source Earth System Digital Twins framework.
* To leverage, improve, and define community standards to promote interoperability.
* It is about delivering professional quality open-source platform that addresses - Agility | Relevancy | Sustainability.
* To enable end-to-end data and computation architecture, and the total cost of ownership.
* To start with system architecture aiming for simple interfaces and information model - from generalization to specialization.
* Must team with Science Champions.
* Successful big data platform needs to be ready for multi-computing, multi-data-center, and multi-agency.

## [Cloud Use Activities and Strategy](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_11.40_SEO_Report.pdf)

David Borges (NASA, SEO) gave a presentation on cloud use activities and strategy. He noted that there are several motivations for a cross-cloud study: a key finding from the Pecora 22 Workshop for the need to capture lessons learned from recent experiences with the three big cloud service providers; a request from GEO to CEOS SEO to inform the evolution of the GEO Common Infrastructure (GCI); improving the user experience; and determining how SEO can support CEOS Interoperability Roadmap.

The SEO has been testing several cloud computing frameworks to understand CEOS data access and technology capabilities. The supported environments and services are Google, Amazon, Sentinel Hub, and Microsoft. Using virtual private server machines, and Sentinel-2 as the data source, early findings are definite differences in performance based on testing of standard options. This is a simple benchmark intended to fortify future refined benchmark methodologies and requires larger statistical experiments over time to truly determine accuracy of benchmarks (daily service volume can vary, new systems can come online, etc.) The following operations were tested for each cloud provider: data loading in an Xarray, calculation of spectral indices (NDVI, NDWI, MNDWI), and plotting. The conclusions are:

* AWS 6x faster data loading than GCP and Azure
* Azure computation 4x faster than AWS and 10x faster than GCP
* Azure plotting 3x faster than GCP and 4x faster than AWS
* Total execution: Azure wins (2x faster than GCP). Disclaimer: this is the first known test of all three major and AWS.

David gave a satellite data summary, and quantified the market share and datasets of cloud providers.

The SEO also conducted a cloud project with Sinergise, testing the Sentinel Hub – Open Data Cube (SH-ODC) environment on CreoDIAS. The demonstration tested confirmed the ability to use ODC on CreoDIAS and utilize the SH-ODC environment for importing Sentinel Hub datasets to be used in ODC applications.

David reported on the Earth Analytics Interoperability Lab (EAIL) - a data and analytics platform that uses AWS Cloud and Open Data Cube. Its advantages are Jupyter Hub, Dask scaling, customized ARD pipelines and GPU processing. There are 59 registered users, and Jonathan Hodge (CSIRO-Chile) is the primary EAIL lead and architect. The SEO is working with CSIRO in 2023 to become trained on EAIL operations to support users. EAIL currently supports one active CEOS project: COAST (Chesapeake Bay study). Other projects interested in using EAIL include: WGCV (DEMIX Cal-Val campaigns), DE-Americas (Caribbean Pilot project), and CEOS Ecosystem Extent Pilot Project. Datasets include: Landsat, Sentinel-2, MODIS, Sentinel-3, Sentinel-1 (CARD4L with RTC), Copernicus DEM, and NASA DEM.

David concluded saying that the SEO plans to participate with the proposed CEOS Interoperability Framework effort. Tom thanked him, noting that the involvement of the SEO will be necessary.

## Discussion on the Vision, Outcomes, Actions, Next Steps

Makoto and Tom led a discussion on the vison, outcomes, actions, and next steps for the CEOS Interoperability Framework task. A roadmap for this task needs to be submitted to the SIT Technical Workshop in September. Below is an organized summary of the discussion:

General

Given the vision of LSI-VC who initiated this project, need to determine what is the final target, the vision, the components of this framework. Are the components covered by existing CEOS work, and by existing WGISS work?

CEOS Interoperability Framework leadership:

Overall responsibility is assigned to WGISS. Makoto suggested that the leadership should be with the WGISS Chair. Ken suggested that the chair team could be the lead, and the tasks be aligned to WGISS key strategies. Another option is to re-activate the Interoperability and Use Interest Group placing the leadership there.

WGISS organization:

If interoperability becomes a key focus area for WGISS in the next three to five years, should WGISS organize its activities around the factors themselves. For example, for the data accessibility factor, does that effort go into the specific interest group, or should the interest group be disbanded to focus on the data accessibility factor. Damiano noted that an increase in groups will result in an increase of meetings, which should be avoided. Esther commented that having an interest group for this effort would provide an advantage.

Interoperability factors:

Several factors have been identified, with interoperability being transversal to all the factors. Each factor needs to have a lead, who will work across all the other groups, and who is responsible for submitting a description and scope of the factor. The factors could be sequenced so they can be prioritized CEOS can start delivering value.

Regarding the Policy factor, Rama (NASA) noted that this encompasses preservation, but felt that preservation should be an explicit factor. One suggestion was that preservation could be thought of as a use case of the framework, or preservation could be incorporated into data accessibility.

Contributions/participation:

Need to determine the resources needed to progress these components and make them coherent.

Esther mentioned that an upcoming UKSA project that could contribute in a significant way.

Data policy is data access, data authenticity, and data availability (versions, number of copies, how long to keep). This may be an area where the SEO is well-suited.

The data accessibility factor can have a JN white paper contribution and the white paper on federations also can be a contribution.

A way forward may be to define a catalogue of use cases. Digital twins could be the first use case, as it fits very well into all the factors. The goal would be to show how the interoperability is set up and show how agency investments can be connected, and gaps can be identified.

How to establish a bridge between cloud platforms and a personal computer/device: The data packaging standards would allow the description of a process that identifies and preserves an image, knowing its provenance. In summary, this is architecture, semantic and syntactic.

Discussion will continue and decisions made at the end of this meeting.

Action WGISS-55-01: Each factor Lead of the CEOS Interoperability Framework to finalize the factor name, coordinator (lead), collaborator(s), description (1-2 paragraphs), scope (included and excluded).

Action WGISS-55-02: WGISS to work with Thomas Huang to consider the digital twins as a use case in the CEOS Interoperability Framework.

# WGISS-WGDisasters Joint Symposium



## [Introduction to CONAE](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_09.10_Int-to-CONAE.pptx)

Laura Frulla (CONAE) gave an overview to CONAE. The main activities are receiving stations operation, satellite monitoring and command, integration and test activities in dedicated laboratories, data reception and storage, data processing and products generation, data distribution, education, capacity building and user training, satellite missions, and access to space.

The Argentine national space plan is a 10-year strategic plan that governs the activities and projects of CONAE since 1994 without interruption. The main objective is to go to space to observe the Earth, making decision-making more efficient in the public and private sectors, and promoting and motivating national development.

EO data processing and products generation cover numerous strategic areas and benefit areas. Education, capacity building and user training is managed from the Mario Gulch Institute and others.

SAOCOM mission objectives were described, and include to operationally integrate SIASGE (Italian-Argentinian Satellites System for Emergencies Management). SAOCOM Catalogue can be navigated freely to identify already existing level 1, level 2, and higher-level products. CONAE also works on the SABIA-Mar mission to measure ocean color.

Argentina’s access to space objective is to provide the country with its own capacity to access outer space with development of launch vehicles capable to inject payloads up to 750 kg in LEO, and providing launch services to Argentine satellites and others from abroad.

## [Overview of WGISS Activities](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_09.25_Overview%20of%20WGISS%20Activities.pptx)

Makoto Natsuisaka (JAXA) gave an overview of WGISS activities. He gave the vision of WGISS and organizational structure, and described the activities and accomplishments of the four interest groups of WGISS: Data Preservation and Stewardship, Data Discovery and Access, Data Interoperability and Use, and Technology Exploration. Makoto listed recent webinars in cooperation with WGCapD. He concluded with a brief description of the Earth Analytics Interoperability Lab.

## [Overview of WGDisasters Activities](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_09.40_%20WG%20Disasters%20and%20WGISS.pptx)

[Hélène de Boissezon](mailto:Helene.DeBoissezon@cnes.fr) (CNES) gave an overview of recent updates in WGDisasters activities, highlighting the areas of synergy between them and WGISS.

The CEOS WGDisasters Work Plan is aligned with the Sendai Framework on Disaster Risk Reduction, acknowledging that EO enables identification of natural hazards, exposure of people, assets and livelihoods, and related vulnerabilities.

WGDisasters priorities are:

* Demonstrators: Continue to demonstrate potential but begin building path to sustainable operations post demonstrator – stronger ties to international stakeholders but also local actors, and increase focus on capacity building in all demonstrator activities.
* Strengthen ties to GEO WGs through increased visibility of WG Disasters activities within GEO.
* Explore linkages to climate related activities, especially through the impact of CC relating to extreme weather events and local impacts of climate (assessment, mitigation, resilience).
* Exploit new technology opportunities, either through new missions, new activities, or new data exploitation techniques.

Pilots and demonstrator projects are a contribution to GEO and include the Wildfire Pilot, the Flood (GEO/LEO/SAR) Pilot, the Landslide Demonstrator, the Volcano Demonstrator, the Seismic Hazards Demonstrator, and the Recovery Observatory Demonstrator

Hélène described the Geohazards supersites, the Kahramanmaras event supersites, the volcano demonstrator, the landslide demonstrator, the Recovery Observatory demonstrator.

Pilot and demonstrator successes include:

* Demonstration of technical feasibility as applied to each thematic area or phase of disaster cycle.
* Working-level, proto-operational uptake of EO-based products from practitioners and end users (e.g., volcano observatories, civil protection agencies, specialized research bodies…)
* Exploration of new partnerships and bridging gaps between different types of partners, including international organizations, governments, research institutes, companies, non-governmental organizations… solution-focused work to improve Disaster Risk Management using EO.
* Demonstration of best practices for combining diverse satellite data to improve understanding and management of risk.
* The new activities envisaged include CC impact/extreme events, synergy with WGClimate, SCO (Space for Climate Observatory), improving disaster risk reduction, and Geodesy for Disasters/GEO4Sendai (TBC).

Links to other CEOS and GEO working groups:

* WGCapD, joint activity to provide training and EO awareness to DRM leaders (El Salvador 2022; Africa and Asia 2023); presentation to WGCapD-12 and UN-Space “Meeting on the identification of needs of MS and UN entities for Capacity Building in the Use of Space Observations”, March 2023
* WG Climate, discussions to collaborate on climate adaptation and impacts at local level (underway)
* WGISS, joint session in person April 2023; discussion on machine learning for EO applied to Disasters and other topics
* SIT, collaboration on sustainability and implication of new partners including private sector EO players (Volcano and Landslide Demos)
* New Space Task Team (NSTT) – contribution to topic 10: “Observation gaps, commercial opportunities to fill CEOS priority requirements for climate, SDGs and disasters”
* Oceans Coordination Group, possible WGDisasters thematic contributions (under review)
* GEO DRR WG, RO one of 1st use cases for EO Risk Toolkit/RiX (UNDRR), Landslides to follow and new GEO 2025 thematic Incubators: WG Disasters working to develop use cases.

## AI/ML

### [Presentation of WGISS AI/ML White Paper](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_0955_Presentation%20of%20WGISS%20AI%20ML%20White%20Paper.pptx)

Yosuke Ikehata (JAXA) discussed the white paper that is being prepared by WGISS which will show a use case of Deep Learning (DL), Machine Learning (ML) and Artificial Intelligence (AI) using EO data. This paper 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 dataset.

The target audience is data users and data providers, and will be made available to CEOS at WGISS-57 in April, 2024.

Andrew noted that the Wildfire use case is the most mature case, but there is also interest in the landslide and volcano. WGDisasters is eager to work with WGISS on the white paper, and to review it.

### [Presentation of WG Disasters Planned Use Case on ML and the Wildfire Sector](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_10.10_Planned%20Use%20Case%20on%20ML%20and%20the%20Wildfire%20Sector%20.pptx)

Mark de Jong (NRCan) gave a presentation for Alan Cantin (NRCan) on the planned use case of ML in the WildfireSat mission. He noted that, in addition to the conventional tools to fight wildfires, EO and ML are also tools as demonstrated by operational wildfire mapping with Sentinel-2.

Just about any machine learning method that has been developed has seen an application or attempted application in the wildfire sector, including supervised learning, unsupervised learning, and agent-based learning. Examples of some commonly used methods in this sector are artificial neural-networks, random forests and decision trees, Bayesian methods, and support vector machines. Mark presented several approaches, such as cloud masking, Unet, waterbody ID.

Alan concluded that it is well established that there are use cases for ML-based algorithms for a wide variety of wildfire applications. ML is not necessarily the solution but it is a solution. Multiple ML-based algorithms for the WildFireSat mission are being investigated to evaluate ML-based vs. rules-based results, and implementation challenges with contracting and NRCan information systems. The primary users are fire agencies in Canada, but they are also exploring with other agencies in Canada such as the atmospheric community, and with the international community.

## [Service Discovery Best Practice Activity](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_11.00_Service_Discovery_Best_Practice.pptx)

Yves Coene (Spacebel for ESA) gave a presentation on the Service Discovery Best Practice activities. He reminded that Version 1.0 was approved and published after WGISS-54 (November 2022), which extends Best Practices for collection and granule discovery (CEOS OpenSearch Best Practices) to “services” (including tools, applications, service endpoints) with intention to “share”. Yves showed an outline of the document, and listed two use cases:

* CEOS OpenSearch Best Practices: “Discover EO Collection” and “Discover EO Granules”
* CEOS Service Discovery Best Practice: “Discover (EO) Service” and “Use Service”

Yves listed five discovery scenarios based on survey of applicable metadata models/elements, and discussed the Best Practice recommendations for the Service metadata model (encoding-agnostic), and service metadata encoding. Yves presented a cross-comparison of relevant metadata models/elements/guidance, and discussed how the proposed Best Practices can be applied to different encodings.

Yves concluded with the following next steps:

* Synergy with “Jupyter Notebooks Best Practice”: include a cross-reference, discuss internal Notebook metadata versus metadata records, include CEOS Open-Source SW and Tools Survey, and “couple” with corresponding collection metadata.
* Share “service/tool” metadata records through IDN, as currently done for “collection” metadata records (DIF10).
* Apply to specific use cases and consider lessons learned in future document updates: e.g., checking the possibility of including the Coastal Observation Application Services and Tools.
* Apply to Notebook discovery.

Makoto added that a goal of this presentation is to help WGDisasters with distributing their services and tools to global users. The contact points are Damiano Guerrucci(ESA) and Yves.

## [Use of EAIL in Flood Pilot and Other Possible Synergies](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_11.15_FloodPilot_EAIL.pptx)

Marcelo Uriburu (CONAE) discussed the use of the CEOS Earth Analytics Interoperability Lab (EAIL) in Flood Pilots (FP) and other possible synergies. He began with the background of the EAIL, adding that within the Flood pilots EAIL accounts have been opened for the different FP subgroups, covering basins and areas of interest across different regions of the globe. Satellite data has been made available in those areas of interest, from different sensors.

Among the flood pilot subgroups, some expressed not needing the Lab at this stage and others expressed interest, especially on Analytics tools for integrated analysis using Jupyter Notebooks.

Other CEOS groups using the EAIL are CEOS Rice Monitoring Community, and CEOS COAST, Coastal Observations Applications Services and Tools Ad Hoc Team. Merrie B. Neely (NOAA) expressed their success with the use of the Lab and has strongly encouraged Flood Pilots to move forward along this path.

Other possible synergies between WGISS and WGDisasters are the use of ML and/or AI for EO Analysis for flood extent and flood depth mapping.

David Borges commented that the SEO is/will be actively working with Jonathan Hodge and Marcelo to reinvigorate the use of the EAIL.

## Outcomes, Actions, Next Steps

Action WGISS-55-03: WGDisasters to work with WGISS Technology Exploration Group to incorporate as use cases into the AI/ML white paper the Wildfire Pilot, and perhaps also the Flood (GEO/LEO/SAR) Pilot, Landslide Demonstrator, Volcano Demonstrator, Seismic Hazards Demonstrator.

Action WGISS-55-04: WGISS to provide WGDisasters updates and/or access to the AI/ML White Paper for their collaboration.

Action WGISS-55-05: WGISS (Yves Coene, Damiano Guerrucci) to invite WGDisasters to provide tools and software developed by WGDisasters so they can be made discoverable to global users through the service discovery.

# Data DISCOVERY and ACCESS



## Session Objectives

Damiano Guerrucci (ESA) introduced the session, outlining the session and its objectives.

## Data Access Authorisation - Toward Federated Access

### [EGI (OpenEO) – Federated Authorisation](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_13.20_EGI(openEO)_Federated%20Authorizathion.pdf)

Valeria Ardizzone (EGI for ESA) gave a presentation on federated authorisation through EGI, whose vision is that all researchers have seamless access to services, resources, and expertise to collaborate and conduct world-class research and innovation.

Valeria discussed two types of authorization: User Attributes-based authorisation with VO/Group membership and role information, GOCDB roles, assurance information, and affiliation with home organisation. Capability-based authorisation with resources a user is allowed to access, and optional list of specific actions the user is entitled to perform.

Valeria discussed three authorization models:

* Centralized Policy Information Point
* Centralized Policy Management and Decision Making
* Centralized Policy Management and Decision Making and Enforcement

Valeria concluded with an explanation of the EGI check-in service, and a graphic of the openEO platform authorization model. She confirmed that this is specific to the openEO platform, but other communities can benefit. The requirement is that the user provide a link to their identity in order to respect the security policy.

### [NASA - Multifactor Authorisation](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_13.20_NASA%20MFA%20Analysis.pptx)

Valerie Dixon (NASA) stated that NASA must use multi-factor authentication to modify data to their systems, but approved Multifactor Authentication (MFA) methods are still somewhat limited. In the Earth Observing System Data and Information System (EOSDIS) ecosystem, current options are Launchpad Authentication using tokens or Personal Identity Verification (PIV) badges, AWS Cognito on a case-by-case basis, and Earthdata Login (EDL), which is identity management only.

Valerie listed the MFA Study considerations, including what and who they must support, and other miscellaneous considerations. She listed the tools currently under evaluation. By WGISS-56, the expect to have completed the analysis and received approvals. For IDN providers this will mean that draftMMT and MMT can merge into one application, and direct API connectivity to CMR.

Doug added that, from a data read perspective, security at NASA is not much of a concern. Security for data write perspective is tight, but may be loosened over time.

For ESA, Damiano said they authenticate users with ESA IdP, but something more sophisticated is needed for others. Another point is the need for authorization at multiple levels of users and policy, including GDPR.

## [GEO-ZARR Software Working Group](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_14.00_GeoZarr%20SWG.pptx)

Christophe Noel (Spacebel for ESA) gave a presentation on the GeoZarr Software Working Group (SWG), which was created in January 2023 with the objective to develop conventions for storing geospatial data in Zarr, a cloud-native, multidimensional data format. He noted that COG is widely recognized as a leading cloud-native data format due to its scalability and efficiency. Christophe discussed the reasoning for choosing Zarr over other formats for a geospatial extension. GeoZarr has support in FedEO and STAC browser.

Christophe outlined the history of the SWG, its members, meeting history, and objectives. GeoZarr convention will follow ZEP-0004 approach.

Next steps are to consider interactions with OGC SWG, and contacts with CF. He invited others to join.

Makoto mentioned using GeoTIFF for two-dimensional data.

## [ARD Standardization in ISO TC211 and OGC](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_14.20_ISO%20TC211.pptx)

Liping Di (ISO TC 211) discussed ARD standardization in ISO TC211 and OGC. He began with the definition of CEOS Analysis Ready Data (CEOS ARD): satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets.

Since CEOS is not a recognized standardization organization, CEOS specifications are not considered as international standards. The need is to expand the ARD to cover other geospatial data to maximize the interoperability between EO and non-EO data.

The approach for standardization of ARD through OGC and ISO is to use CEOS ARD specifications as the base, building on existing OGC and ISO standards to achieve broader community consensus. This is a joint ISO and OGC standard development effort that will go through formal standard development process and be published as ISO and OGC standards. The CEOS ARD Oversight Group, CEOS WGISS, OGC, and ISO TC 211 have all agreed on the approach.

OGC ARD SWG and ISO TC 211 ARD project team will work together to develop ARD standards. The final version of SWG charter was presented at the Plenary of 125th OGC member meeting in Frascati, Italy, and was unanimously approved to forward the Charter to P-members for voting.

ARD activities at other OGC initiatives include OGC Climate Resilience Pilot, the OGC Disaster 2023 Pilot, and the OGC Testbed 19.

The first ISO ARD project team meeting will be held in South Korea on May 16, 2023

ISO has designed ARD series of standard to be ISO 19176: ISO 19176-1: Geographic information — Analysis Ready Data — Part 1: Framework and Fundamentals.

The initial working draft is being prepared as are the meeting materials for the project team meetings on May 16 and June 7. The main issue is to obtain funding support to key persons are needed. Since June 2022, no funding has been provided to project leaders on ISO/OGC ARD work despite oral commitment from multiple organizations.

All work currently is volunteer basis. A broad community consensus is needed for the standard to be successful.

## [IDN and CWIC Update](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_14.35_IDN_CWIC_Report.pdf)

Michael Morahan (NASA) gave an update on IDN and CWIC. He began with discussion of the IDN and CWIC portals, metrics, and recent collaborations with ESA/FedEO, ISRO, NOAA-NCEI, CCMEO, and USGS.

Michael reported that Unified Metadata Models for Service (UMM-S) / Tool (UMM-T) records in CMR can be searched using CMR API. Users can search for CEOS tools by using ShortName or keywords.

The Direct Distribution Information (DDI) main element allows data providers to provide users information on getting direct access to data products that are stored in the AWS-S3 buckets when they are initially looking at a collection. The end users get information such as the S3 credential end point, a credential documentation URL, and bucket prefix names, and an AWS region.

There are multiple sub-elements that comprise DDI. Each of these elements describes the information that is necessary to pull out data products that are stored in the AWS cloud using a S3 URL that is located inside each granule's (data product file's) metadata.

Adding the CMR OSDD tag is no longer needed to implement OpenSearch collection-specific, granule searches for CEOS providers.

Michael described UMM collections and services updates, and announced that NASA CMR has improved how to handle End-user license agreements (EULAs) in CMR.

There have been ten GCMD Keyword releases since WGISS-54 with various additions to Science Keywords, Platforms, Instruments, Providers, Projects, Services Keywords, and Locations.

Michael concluded with details of IDN metrics.

Tom asked if most of the providers are familiar with adding/removing data? Michael replied that a lot is coming from FedEO and there is great communication between the two teams and also good communication with ISRO and JAXA. With USGS it just works.

Damiano congratulated Michael and Yves with ongoing communication to keep everything working.

## [FedEO Update](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_14.50_FedEO_Report.pptx)

Yves Coene (Spacebel for ESA) gave a report on FedEO activities. He noted that FedEO has been in the CEOS domain since 20 September 2021.

The FedEO Metadata Mediator continuously updates with an automatic procedure available on the FedEO reference environment, and on the operational environment at ESA since Oct 2019. Collections are being harvested automatically by IDN, e.g., ESA, DLR, VITO, ESA CCI. Further, JAXA G-Portal 1017 is ready to ingest in IDN, and additional VITO collections.

FedEO metrics evolution: granules have risen from 123.8 to 148 million granules (+20%) since WGISS-54 and ESA (EOCAT) includes172 collections with 28.8 million granules

In terms of software changes, ISO19139-2 and DIF-10 have been improved (DOI and Associated DOI encoded differently in ISO metadata).

Yves discussed the next generation of FedEO: FedEO STAC interface is being enhanced implementing existing API + existing API extensions.

Data Management and Stewardship Maturity Matrix may be included in the catalogue for collections (if available), and preliminary encoding with ISO19157-2 is underway.

Makoto thanked JAXA and NASA for helping with the ingestion of the JAXA, now in the operational server.

## [Metrics for Cloud-based Data Use](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_15.20_Metrics%20for%20cloud-based%20data%20use.pptx)

Doug Newman (NASA) and Min Wong (NASA) gave a presentation on metrics for cloud-based data use. Doug highlighted that Cloud access patterns are evolving from traditional access patterns. This may create perceived discontinuities between traditional and cloud access metrics that need to be understood before drawing conclusions.

Doug described the evolution of data access. It began by getting the whole file(s) into the local system and maybe use some of it, to expansion of metrics beyond data volume into user dimensions, and then getting some of the file into the local system, using most of it. Data access has continued to leveraging the cloud to provide necessary compute, and improving cloud-based data access user experience.

The technologies presented are the Elk Stack and Open Telemetry. Minnie continued that the NASA-ESDIS Metrics System (EMS) that provides charts of monthly distribution trend, volume, and product counts in the past year. For NASA cloud monitoring they use an ELK Stack in AWS to measure egress. The challenges are understanding new access patterns, and comparing ‘apples to apples’ with respect to traditional versus cloud-optimized data access.

New measures of success include analysis-in-place, a viable pattern for the community. NASA’s cloud architecture is ‘cost-optimized,’ and usage patterns point to architectural changes.

Iolanda requested that Doug Newman check the Best Practices package that DSIG will circulate for review to ensure the cloud metrics is considered.

Discussion: This is a work in progress, with ongoing work on further refining metrics and finding ways to present the data to a non-technical audience. These considerations were raised:

* The tool APPEARS may provide a reduction ratio, comparing prior years to present for the same value.
* How the data is moving: from private hosting to public cloud and other different types of environments.
* Volume of data that the service asks for or what the user requests.
* Malicious users who touch the file to make it look like access.

## [Service Discovery: OSS Tools Metadata Ingestion Use Cases – FedEO](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_15.40_Service_Discovery_OSS.pptx)

Yves Coene (Spacebel for ESA) gave a presentation on OSS tools metadata ingestion use cases in FedEO service discovery. He discussed the following topics of the FedEO implementation evolution:

* FedEO RESTful API
* FedEO evolution implementation topics and examples
* STAC API and STAC
* Open-Source software and tools survey
* Notebook discovery
* FedEO API supported data formats
* FedEO API SPARQL prototype

Next steps and future work include:

* STAC
  + Covers granule (item) and collection metadata, no (item) encoding/support for service/tool metadata.
  + Limited support for “coupling” collections/granules with service endpoints
  + STAC API might nevertheless be used for discovery (as “separate”) from STAC Item specification
* ESA SKOS thesaurus extension for Tools/Services with cross-reference to GCMD
* Subset to become operational as EOCAT evolution
* Cloud-Native Demonstrator, including coupling of EOCAT collections   
  with DataCube (service) metadata
* Improve CEOS Service Discovery and Metadata compliance
* Include/host Notebook metadata, possibly extracted from Notebook internal metadata (CEOS Notebook Best Practice)
* Support UMM-JSON export format to IDN
* Useful for WGDisasters?

Iolanda shared this link to the Open-Source Software Inventory, which may be a duplication: <http://ec2-3-208-162-171.compute-1.amazonaws.com/ossi/browse>. Since the tool is now visible, WGISS should ensure the correct content is on the operational server of FedEO so it can be accessed by users and then have the access to the old inventory removed from the CEOS site. This should also be communicated to the users.

There may be tools to ingest in the same way as the collections; currently it is manual, but there will be a UMM.

Action WGISS-55-06: DSIG/SLT to confirm that content of the [Open-Source Software page](http://ec2-3-208-162-171.compute-1.amazonaws.com/ossi/browse)  on the CEOS website has been correctly ingested in the FedEO operational system that is accessible to end users.

Action WGISS-55-07: DSIG to ask SEO to remove from the CEOS website the [Open-Source Software page](http://ec2-3-208-162-171.compute-1.amazonaws.com/ossi/browse), upon completion of previous action.

Action WGISS-55-08: SLT to define a procedure for registration of services/tools into IDN/FedEO (direct registration by users or through circulation of Open-Source Software spreadsheet). SLT to inform WGISS-All on how to register their services and tools.

## Outcomes, Actions, Next Steps

Damiano Guerrucci (ESA) led a discussion on outcomes, resulting in this action:

Action WGISS-55-09: SLT to examine the emerging needs and possible solutions for user authentication and authorisation in the new scenario for data access. Provide a status report at WGISS-56.

# Data PRESERVATION and STEWARDSHIP



## Session Objectives

Mirko Albani (ESA) outlined the session and its objectives.

## [Archive Technologies Evolution White Paper](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_09.20_Archive%20Technologies%20Evolution%20White%20Paper.pptx)

Daniele Iozzino (RHEA for ESA) gave a presentation on the white paper on archive technologies evolution. A snapshot of the surveys on Long Term Archive Technical Meeting sessions addressed:

* Archive Infrastructure and technology
* Monitoring tools
* Archive volume, missions, coverage
* Archiving flows and processes
* Data format/packaging for long term archive
* Management of the relevant associated information, documentation, software, and knowledge (if managed)
* Trends and Future Plans, Main Issues and Challenges

The main challenge of the archive is the sheer volume of data, the copious number of files, the diversity of formats and the yearly increase of data produced. For example:

* Lots of I/O affects performances due to the iNodes being constantly updated
* Performances are still scarce for small files even with iNodes on separate partition/RAID
* Media failures recovery takes time and uses resources (drives, storage)
* Lots of false media failures alarms (caused by Network or robots)
* Current Archive manager restricts since it is vendor (and OS) locked
* Support tickets requires lots of data collection even for obvious failures
* Support tickets often result in requests to update firmware
* Firmware updates are complicated time consuming
* Current HSM solution is cheap compared to alternatives
* Archive manager migration complicated, time consuming, expensive
* Longevity of technology has side-effects (maintenance costs, obsolescence of HW)
* Tape prices are not predictable. T10K tapes prices doubled when ORACLE declared the end of life of the tape family. LTO-8 tape shortage in mid-2019 due to patent infringement battle meant LTO-8 tapes were unavailable for a long period of time
* With the end of some tape families and HSM archive managers, it is feared that the main vendors are moving away from tape archive solutions
* They are ready for lots of data, less ready for billions of files
* Extracting data from tapes is slow and complicated

Due to lots of I/O necessary to update iNodes, performances when archiving small files are scarce. The archive has been updated to cope with such limitation and the iNodes have a dedicated partition on two SSD disks in RAID1.

Performances are still scarce and were recently challenged by the bulk ingestion of ESAC Scientific Archive where such a copious number of files was not expected. Consequences of such high number of files have been assessed. Other issues are media failures, and vendor-locked solutions.

The advantages of tape archive include: easy to expand, power efficient, affordable tape evolution on par with disk evolution, malicious encryption and impossible deletion.

The archive at ESRIN is constantly enhanced to facilitate the archival process by using technological innovation and practices. Green computing technologies are favorable to save energy, thus making the archive and its hardware environmentally friendly.

A study is underway to help on a possible evolution of ESA ESRIN: Archive with current or future technologies, review of hardware, software, services, external tools, environment-friendly, correct preservation guarantee. Key findings include:

* There are solutions to keep using current ESA/ESRIN storage technologies. You do not need to change the entire infrastructure because of HSM end of life
* Hardware providers are developing hot, warm, and cold storage solutions that can complement the current infrastructure and, defining a roadmap, could substitute it in the long-term
* Some WORM storage options are already available or being developed that are based on nonmagnetic media, like optical disks or crystal based. There are some that could help ESA on the long-term perspective without needing to migrate as the media lasts for a longer time
* More and more, organizations are moving to cloud-based services because of their flexibility and cost effectiveness in the short-term. However, the step to cloud needs to be planned because long-term costs and vendor locking could be an issue in the future
* Complementary to pure cloud solutions, the hybrid-cloud approach suggestable. A mix between public access on the cloud plus cold on-premise storage would benefit in the long-term to address purecloud flaws
* The preservation software market is very active and growing. In addition to the classic providers in this area, some new players are showing interesting solutions also with a very open-minded approach to avoid extra costs or cloud locking. SaaS is a trend on this market but private cloud solutions are also available
* Security and sustainability are also worries in this market. Many providers are showing some environmentally friendly messages but it is not clear how much are they real or just green-washing
* OAIS is the standard for preservation and needs to be followed. In addition, other standards could improve ESA/ESRIN archive operations in different directions
* Big cloud providers are charging for requests or operations which makes it almost impossible to make a budget in advance
* Some European cloud providers are not charging for the recovery or the model is more manageable

Daniele displayed an outline of the upgrade roadmap, and listed the following conclusions:

* After analyzing the market, the current situation of the Archive and the reference guidelines the best solution for updating the Archive is a hybrid cloud model, which integrates the current tape storage infrastructure with a flexible model that facilitates access and improves its online capabilities.
* The current Tape Library work-life has been considered sufficient to cover the needs of the coming years and therefore, the recommendation is to continue using it for some years now. The archive manager shall be replaced by any of the identified alternatives or by an integration into cloud services.
* One understated advantage of continuing to use tape-based archives is the always increasing security concerns. Malicious encryption or deletion of the data is not possible or highly complicated. Moreover, to reduce energy consumption and carbon emissions using tape storage instead of hard disk is an effective solution.
* ESA is defining a roadmap for the evolution of its ESRIN archive. The roadmap is the first step in the direction of assuring seamless archival, extraction, valorisation, and preservation of the always increasing ESA and Third-Party Missions (TPM) EO data records and associated information.

The final goal of this presentation is to approve the proposed Table of Content of the CEOS WGISS deliverable: Archive Technology Evolution White Paper, whose first draft is expected by the end of May:

* Introduction and Scope
* Current Archive Solutions in EO
* Main Issues and Challenges
* Archiving technology trends
* Recommendations and way forward

Nitant asked if they are considering object storage as a mechanism for archival. Daniele replied that they are indeed, and that all their software can interface.

Yousuke asked if they use hybrid cloud archiving, and if the white paper will consider cloud/hybrid egress. Daniele replied that yes, what they are moving to cloud is not cold storage, but the processing, a hybrid solution. Mirko said the hybrid solution to put the software on cloud and leave the storage on-premise, may eventually lead to data moved to cloud, but must consider egress cost and data availability. The idea of the white paper is not to provide a solution but to present the current solutions, their issues and challenges, and what are the archiving technology trends, pros and cons. Recommendations will allow agencies to decide among the various solutions.

Homero asked how much time to shift from one reel to another. Daniele replied that it is few seconds and up to maximum 2.5 minutes. The data is written on disk storage ,then written to tapes. 18 TB per day to write, and 20 TB per extraction.

WP draft will be released in May.

## [DSIG Best Practices Refreshment](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_09.40_WGISS%20DSIG%20Best%20Practices%20Refreshment.pptx)

Iolanda Maggio (RHEA for ESA) outlined the goals with refreshing the Data Stewardship Interest Group the Best Practice documents:

* New terms introduced by the rapid technology evolution
* Updates of the document references, considering the BP created in different periods
* Harmonising statements, considering the full vision of the BPs in WGISS-DSIG landscape
* Adding/updating, when needed, use cases, recommendations, and guidelines

This effort closes the following actions:

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.

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.

For each document Iolanda detailed the updates performed and decisions made on the content:

1. Preview Image Principle: content verified. Contains a lot of information on browse guidelines. Content verified.
2. Data Management Statement: Content verified and document removed.
3. Long Term Preservation of EO Space Data Preservation Workflow: Content verified and document updated.
4. Preservation Guidelines: Content verified and document updated.
5. Preservation Dataset Content: Content verified and document updated.
6. Persistent Identifier Best Practice: Content verified and document updated.
7. Technical Content and Associated Information Preservation Best Practice: Content verified and document updated. Should this be a formal White Paper for WGISS?
8. Measuring Earth Observation Data Usage Best Practice: Content verified and document updated.
9. In-situ Data Analysis: Should this be a formal White Paper for WGISS? For the nature of the in-situ data all Best Practice documents are extended to include also this concept
10. Cloud Metrics Data Analysis: Content verified and document updated.

White Papers and Reports:

* [Long-Term Archive Strategies](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Data-Long-Term-Archive-Strategies.pdf) 🡪 (2011) It could be reviewed and included in the Archiving Technology Evolution Report
* [Data Preservation Techniques](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Data_Preservation_Techniques.pdf) 🡪 (2011) Old White paper already covered by our Best Practices
* [Data Lifecycle Models and Concepts](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Data-Lifecycle-Models-And-Concepts.pdf) 🡪 (2012) Theoretical description of Data Model. Superseded by Reference Model?
* [Browse Survey 1997-2010](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Browse-Survey-1997-2010.pdf) 🡪 It contains the “results of a March, 1997 survey of some CEOS agencies regarding the details of their production of AVHRR browse.” and the Browse Guidelines “This document is intended to assist data providers, including designers, developers and operators of Earth observation data systems to understand how a browse service might be made available to a wider audience.
* [Guidelines for GIS-Ready Products](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Guidelines-For-GIS-Ready-Products.pdf) 🡪(2011) It is a Draft for Comment
* [Browse Guidelines Document (version 2)](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_DSIG_Browse-Guidelines-Document-v2_Aug2013.pdf) 🡪 Link broken
* [Offline Media Trade Study](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Reports/WGISS_DSIG_Archive-Media-Trade-Study_Aug2012.pdf) 🡪 (2012) ARCHIVE AND RECORDS MANAGEMENT
* [Data Purge Alert White Paper](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/WGISS_DSIG_Data%20Purge%20Alert_WP.pdf) 🡪 No review needed
* [WGISS Data Management and Stewardship Maturity Matrix White Paper](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix.pdf) 🡪 Already Reviewed and ready
* [WGISS Data Management and Stewardship Maturity Matrix](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix.pdf) 🡪 Already Reviewed and ready
* [WGISS Data Stewardship Reference Model White Paper](https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Data%20Stewardship%20Reference%20Model%20White%20Paper.pdf) 🡪 No review needed

Tom suggested adding a potential review date on all the documents WGISS releases, and that retired documents not be deleted, but rather marked as retired.

A discussion on in-situ data noted that a study has been conducted and a report produced with the following content:

* Definition of the in-situ data
* Categories of in-situ data
* Sources of in-situ data (Ground-based, Marine-based, Air-based
* Benefit of in-situ data
* Current in situ data usage in Earth observation
* Challenges facing effective use of in situ data (Quality and consistency and Reliability and accessibility)

Ken’s opinion was that, since the documents were already updated to say that these BPs are applicable to in-situ data, all that is necessary is to release a short report with justification that the BPs apply to in-situ. Others thought that, since the research was done, it only makes sense to write it up. Nitant recommended sending the report to WGCV for comments.

Regarding Technical Content and Associated Information Preservation Best Practice: It is worth formalizing since the work has been done, even though this is not WGISS domain. Conclusion was to review for completeness and timelines, and prepare a Technical Note.

Yves recommended that WGISS publications clarify their type (white paper, technical note, best practice).

Action WGISS-55-10: DSIG to prepare a white paper on in-situ data based on the study conducted, and submit it to WGCV for review. Review of the content should be completed by the end of May, and the document should be circulated by end of June if no updates are needed.

Action WGISS-55-11: DSIG to prepare a white paper on software preservation based on the study conducted. Complete the review of the content by end of May, and circulate by end of June if no updates are needed.

Action WGISS-55-12: DSIG to circulate the package of WGISS updated White Papers and Reports to WGISS-All to decide if they should be updated or retired by the end of April. The WGISS website should be updated accordingly by the end of May. DSIG should also clearly state the issue date, suggested retirement date, and the type of document (White Paper, technical note, Best Practice). The list should be submitted to the VC leads for their consideration.

Action WGISS-55-13: Makoto Natsuisaka to ask the SEO and CEO to provide a method of governance on the process for submission and “publication” of CEOS documents (White Papers, Best Practices, Technical Documents, and deliverables from the Work Plan).

## [Copernicus Sentinels Ground Segment Architecture](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_10.40_Copernicus%20Sentinels%20Ground%20Segment%20Architecture.pptx)

Razvan Cosac (ESA) gave a presentation on Copernicus Sentinels ground segment architecture. Razvan described the legacy Copernicus ground segment architecture, its evolution, and its principles. He further discussed its more recent transformation, the current ground segment setup, and the long-term archive and data access services.

The Copernicus Data Space Ecosystem will continuously expand over the upcoming months. The release occurred in January, with initiation of user registrations, initial data offering, browser, and catalogue APIs. In April, STAC and S3 were completed, and processing, traceability and on-demand production APIs were released. Key public service capabilities and data offering will be available by July 2023, including the full archive of Sentinel missions, complementary open datasets, access, to commercial data, extended Sentinel Hub APIs, OpenEO, Jupyter Lab, and Marketplace. Finally, in November, Sentinel engineering and auxiliary data, Copernicus Contributing Missions, streamlined data access of federated datasets will be possible.

Ken asked about the open ecosystem on the public cloud. Razvan replied that it would be operated on two clouds, with users able to deploy their own services with their own data, for their own community. The public cloud is provided by industry in Europe.

Mirko asked if the absence of level 1 data in the archive is to save space? Razvan replied that the idea is that level 0 is the baseline from which the higher levels can be generated; this is a tradeoff between resilience and cost. They do archive a small amount of level 1 that is not made available.

## Session on AVHRR Data Archives

### [AVHRR Data Recovery Project](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_11.05_AVHRR%20International%20Cooperation%20Activities.pptx)

Mirko Albani (ESA) gave a presentation on the AVHRR international cooperation activities and partners. The objective is to:

* Unfold and make accessible 1km AVHRR data from regional archives (possibly open and free)
* Transcribe unique AVHRR data from heritage media
* Identify a common format for AVHRR Level-1b and Level-1c data and pursue (re)processing from AVHRR data owners/holders and data accessibility
* Facilitate data discovery through the WGISS Connected Data Assets Infrastructure

Since WGISS-54 an inventory of existing national/regional HRPT and LAC data archives under consolidation. Version 2 was updated and presented/circulated at WGISS#-5 and version 3 will be produced at end 2023. A set of high priority datasets were identified for further action: South Africa (SANSA), Argentina (GiDyC-Servicio Meteorológico Nacional), Brazil (INPE).

Identification of high priority heritage media to be transcribed and approaches/opportunities:

* Pilot activity started by ESA with a commercial company to assess feasibility of transcribing three types of Optical Disks containing unique AVHRR data: successful transcription of the provided samples; negotiations ongoing for full transcription of 300 disks from UK university of Reading plus additional available at ESA.
* Successful transcription of 2 tapes containing data acquired in Malindi (Kenya) and received from ASI/ University of Rome. Transcription of several hundred DLT tapes will be carried out in 2023.

Data discoverability and reprocessing will be addressed at a later stage.

Mirko concluded with details on multiple AVHRR regional archives, and process for transcription of optical disk and DLT tape data.

Ken mentioned a similar survey with similar goals that never came to good conclusion. He reminded Mirko to review the material on this to see if it is useful, as it might provide other AVHRR archives from around the world that could be included.

Action WGISS-55-14: DSIG to organize a session on AVHRR at WGISS-56.

### [AVHRR Data: European Coverage](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_11.05_AVHRR_Europe.pptx)

Mirko Albani (ESA) and Stefan Wunderle (UniBern) discussed the project for 40 years of European AVHRR-LAC data – compiled, consolidated and now accessible via ESA dissemination service. The primary motivation is to make the AVHRR historical data accessible to the public and keep the data alive for an unlimited time.

Mirko described the AVHRR Curation Project, listing the steps implemented to compile an open accessible AVHRR dataset via ESA dissemination service.

A homogenous and consolidated AVHRR LAC time series (1981 – 2021) is now available via ESA dissemination service, with more than 250000 AVHRR data (level 1b) covering Europe ready to be used. Approximately 55000 CEOS Sharp-1 segments were rescued, re-processed in a consistent way (EO-SIP) and are accessible via ESA dissemination service. Software and processing procedure developed at University of Bern is installed and tested at ESA facilities. The next step is the generation of Level 1c data (calibrated and geocoded) for a better service to support communities without the needed expertise in AVHRR processing.

The future outlook includes:

* Filling of AVHRR level1b archive with LAC data until the end of AVHRR sensor (NOAA, MetOp) approximately in 2025.
* Integrate global AVHRR LAC data of the pre-MODIS era (e.g., 1992-1999; more than 30000 datasets); pursue rescue activities for local archives around the world.
* On the way to a FCDR: more precise uncertainty characterization (FIDUCEO) and inclusion/alignment with Sentinel-3 and other medium resolution satellite data under consideration.
* Data Usage: process AVHRR data as much as possible in alignment with CEOS ARD specifications and provide Analysis-Ready-Data (ARD) for climate modelling user community (CMUG).

### [AVHRR Data: North American Coverage](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_11.05_AVHRR%20Data%20North%20American%20Coverage.pptx)

Ken Casey (NOAA) gave a presentation on AVHRR data availability at NOAA. The NOAA archives AVHRR LAC, HRPT, and GAC data at the National Centers for Environmental Information (NCEI) are using the CLASS system. AVHRR data of various forms are archived by many other organizations as well. Ken focused on the AVHRR LAC and HRPT data in the NCEI/CLASS archive, and some information provided by NASA. NCEI did an inventory analysis based on filenames; further analysis on actual geographic extent would take more effort.

Ken reported that NOAA assessed its 1 km AVHRR data available in its Comprehensive Large Array-data Stewardship System (CLASS), conducting an initial assessment of both HRPT and LAC data. The analysis is focused on historical data, beginning with 1978 and ending with 2010. He displayed HRPT + LAC Basic Statistics in CLASS, and discussed spatial and temporal coverage.

Mirko asked for more detail on access to certain data from the stations in the US. Ken will follow up with Mirko.

Homero asked about scenes captured on the same day in different stations that could be duplicates/overlaps. Is there an urgency to gather a timeline? Ken said that such a project is not planned yet; rather they are trying to steward the data to make it useful in such a way that access would not require a lot of effort.

## [Introduction of DOIs to the JAXA G-Portal Products](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_13.00_Assignment%20of%20DOI%20Numbers%20to%20JAXA%20Earth%20Observation%20Satellite%20Data.pptx)

Yosuke Ikehata (JAXA) gave a presentation on the assignment of DOI numbers to the JAXA EO satellite data. He discussed the landing page.

Yousuke reported that JAXA decided to assign a DOI number to a collection, not to individual granules and JAXA does not give a new DOI number when new product versions are made. The DOI numbers were registered through Japan Link Center (JaLC) and to DataCite via JaLC.

Yousuke concluded with the compatibility with the CEOS PID best practice of this effort.

## [Data Replicas (Authenticity and Integrity) in the Cloud Use Case](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_13.10_Data%20integrity%20and%20Authenticity%20on%20Cloud.pptx)

Mirko Albani (ESA) gave a presentation on a use case for data replicas (authenticity and integrity) in the Cloud. Based on use case scenarios from NASA presentation at WGISS-54. In the ideal case all the authoritative data required is available on the same cloud platform, and the user only requires access to the cloud storage and notification mechanism. In the worst case, different data products are unavailable on the cloud platform, and the burden falls on the user to replicate the archive from different authoritative sources requiring the user to have a uniform scalable data access mechanism.

In the common case, some data products may be available on the selected cloud platform, but their provenance is unknown, and the user now must not only move the unavailable data but also verify the authenticity of available data. In addition to a uniform approach for data access for the missing data, the user needs a uniform approach to verify the authenticity, completeness, correctness, and consistency of the data available on the cloud.

Mirko described several human and technical solutions, noting that each solution has pros and cons.

Mirko reported that ESA and NASA recently performed a successful large-scale data transfer using a cloud infrastructure to transfer Sentinel-2 data between the two agencies. ESA was able to provide access to NASA for the S2L1C archive between November 28, 2015, to September 30, 2020. This archive has almost 11 million products with a total size transfer of 5.7 PB, divided into 89 buckets. NASA initiated this archive to be transferred from OVHCloud-hosted cold storage (OpenStack Swift based) to NASA-controlled AWS S3 storage in the US West region. The final solution is based on open-source Apache Airflow and Apache Airavata MFT.

The transfer was optimized to the network bandwidth. NASA ran 20 parallel agents to pull the data at a rate of 20 Gbps. The 5.7PB archive was transferred in 27 calendar days. An integrity check post-processing workflow was implemented using a digest java utility to validate the merged objects against the source checksum and re-transferred the mismatched files. Data transfer activity and results are detailed in a paper submitted to IEEE.

Tom commented that a lot of the solutions deal with entire files or scenes. An additional challenge is when a user applies a service to a piece of the file.

## [Authenticating Data Replicas](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_13.30_%20data%20replicas%20in%20the%20cloud.pptx)

Doug Newman (NASA) and Michael Morahan (NASA) gave a presentation on authenticating data replicas in the Cloud. They described the options to check integrity: Checksum in metadata, Data Upload and Verify, and Blockchain provenance. They also described technologies for replication: AWS Cross-Region Replication (CRR), Snowflake, and RClone. Two use cases were presented:

Use Case 1: as-is archive replication, where a data provider may want to replicate a data archive in order to abide by data sovereignty laws, abide by data localization laws, meet compliance requirements, minimize data access latency, and maximize collaboration within the earth science community

Use Case 2: value-added archive replication, where a data provider may want to replicate a data archive in a different format in order to allow users to leverage cloud-optimized access techniques such as re-process as Zarr, Cloud-optimized GeoTIFF (COG). Or the data provider may want to provide an ‘archive of convenience’, replicating/combining portions of data for a specific scientific use case and provide additional assets.

One example is Pangeo Forge production: grass-roots analysis ready, cloud-optimized (ARCO) datasets have been popping up in various cloud environments recently.

Another example is Pangeo discovery, with provenance being achieved by being completely transparent. The agency stamp of approval is on the process, not the data.

Esther asked if this method can be used with third party files? The third party would have to be encouraged to use a process like this.

# TECHNOLOGY EXPLORATION



## [Session Objectives](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_14.00_TE_SessionObjectives.pptx)

Yosuke Ikehata (JAXA) listed the primary goals of the Technology Exploration Interest Group:

* Give CEOS/WGISS members opportunities to collaborate and discuss present and future technology solutions.
* Research technologies that can help the Earth observation community be flexible and adaptable in their IT infrastructure.
* Help facilitate CEOS/WGISS understanding of all generations of technology and support the implementation of both legacy and leading-edge technologies into Earth observation data systems.
* Promote technologies in CEOS/WGISS that prove beneficial to the Earth observation community.

The expected outcomes of the session are:

* Contribution to the CEOS strategic goal: Promote Data Democracy by improving access to and use of CEOS agency data.
* Contribution to the CEOS Interoperability Framework Initiatives.
* Contribution to
* capacity building through EOTEC DevNet.

## [Status of Jupyter Notebooks Best Practice](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_14.10_JupyterBP.pptx)

Esther Conway (UKSA/NCEO) gave an update of the WGISS Jupyter Notebooks Best Practice (BP). She displayed an outline of the document draft, and showed the agenda of the Jupyter Notebooks Day – October 21, 2022. Esther listed the target communities that will benefit from the BP. The objectives are:

* Clear understanding of 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.
* Encourage the development of good workflow and structure within notebooks along with quality documentation. Support their reuse and adaptation by new users
* Support discovery of relevant notebooks in terms of dataset, domain, application/function and skill level
* Delaying technical obsolescence and ensuring longevity of relevant notebooks
* Maintaining the quality of archive by timely removal of redundant notebooks
* Supporting interoperability of notebooks on different platforms
* Lowering barrier for EO data exploitation and raising technical skill level

Esther noted that an access - metadata workshop, and a dependency workshop are needed. Also needed is the development of dependency exemplars.

Esther discussed citation of input data and data access, referencing the CEOS persistent identifiers document. There is a need to create a bidirectional link between notebooks and the archived data, with linkages into the IDN or other portals – (Is this an appropriate approach?) Version control, preservation, archival, and software licensing should be addressed.

Once the initial BP is off the ground, an event is needed–role and activities are TBD.

The outstanding activities/issues are:

* Is a registry to help new users navigate the landscape needed?
* How to help full range of agencies and developing countries engage?
* Resolve access metadata – workshop?
* Resolve technical dependencies – workshop?
* Draft the Best Practice (delay 6 months in work plan)
* Engage WGCapD
* Hackathon to generate exemplars for CEOS
* Registry/Repository discussion

Esther commented that she is trying to understand the best approach going forward. She recommends two workshops this year should to drive out the final issues. The outline is defined, but a clear identification of who to invite is needed, including the core group of people with mature examples.

The issue of software licensing and mechanisms for citation was discussed, and Yousuke requested that the exemplars be presented at WGISS-56. The long-term vision for the BP document is that an organization should be able to find and use it. It is expected that WGCapD will provide training.

## Artificial Intelligence/Machine Learning (AI/ML)

### [Proposal for AI/ML Survey in WGISS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_14.25_Proposal%20for%20AI%20ML%20Survey%20in%20WGISS.pptx)

Yosuke Ikehata (JAXA) described the background and scope of the proposed AI/ML survey white paper, which is to show a use case of Deep Learning (DL), Machine Learning (ML) and Artificial Intelligence (AI) using EO data. Some CEOS agencies provide DL/ML/AI data and platforms and discussed them at WGISS meetings, but there is no summary documentation. 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 dataset.

The audience for the paper is data users and data providers. To proceed with the paper, what is needed is:

* Use cases for the white paper (NASA, NOAA, ESA, INPE have already introduced their studies, but they need to be written/submitted).
* A lead editor for white paper.
* Writers and reviewers of the paper as it proceeds.

The draft will be shared at WGIS-56, and published at WGISS-57, including posting on the WGISS website.

Tom asked if CEOS has guidance on white papers and BPs? Is there information on how much these documents are accessed? Doug confirmed that the BP for OpenSearch was foundational and was well used, but getting the word out about documents is unclear. Marie-Claire noted that the communications team lead by SEO can help with promoting and targeting these, and there are no guidelines for a white paper. Mirko added that his team has differentiated these documents by considering a BP to be a set of recommendations with associated definitions, and a white paper is a technical document discussing evolutions, trends, needs, challenges, solutions.

Action WGISS-55-15: Michelle Piepgrass to find out number of downloads from CEOS website (Best Practice, White Papers, presentations) and provide the results to WGISS Exec.

### [NOAA Steps to the Geoverse](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_15.10_NOAA%20Steps%20to%20the%20Geoverse.pdf)

Ryan Berkheimer (NOAA) discussed NOAA’s steps to the Geoverse. The UN-GGIM in the July 2022 Discussion Paper on the Geoverse recognizes the significance of a distributed and machine dominated world. It is understood that disruptive circumstances require disruptive solutions.

* NOAA is currently developing a holistic capability to enable the Geoverse System of Systems (SoS) step change, with a Cloud-based agile reference architecture for evolutionary governance and open innovation, and iterative inclusion and improvement of existing systems.
* The capability is being driven by many requirements and considerations. Web 3.0 standards vision for full semantic interoperability 🡪 achieving the original vision of the WWW. Full definition and process provenance for trust and scientific reproducibility 🡪 treating process as data. Support of federated, multi-owner earth systems digital twins and sensor networks.
* The capability is being pursued through many intraagency and interagency partnerships: Interagency Study Groups, NSF - Open Knowledge Network, Strategic, Industry and Standards.

Ryan discussed a federated framework of universally useful understanding of the following:

Critical Understanding #1: Process as a Dataset

Critical Understanding #2: Semantic Interoperability

Critical Understanding #3: Process as a Digital Thread

Critical Understanding #4: Legal Interoperability

NOAA continues to socialize and coordinate integration with the framework in a controlled environment (the NESDIS Common Cloud Framework), refining implementation details, and assessing and improving performance:

* Partnership contexts are refining and validating higher order interoperability and functional requirements
* Continuously mapping back to Geoverse goals and milestones to get a sense of gaps and timelines
* Tackling problem of legal interoperability using widespread collaborations

### [Plymouth Marine Lab Work](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_15.35_PML_AI_Talk.pdf)

David Moffat (PML for UKSA) discussed machine learning in marine research at the Plymouth Marine Lab (PML). He listed the projects underway, including:

* Few-Shot learning for the identification of FlowCam plankton images
* Detection and Classification of Floating Plastic Litter Using a Vessel-Mounted Video Camera and Deep Learning
* Pacific Oysters Detection Using Drones
* Detection of ship tracks
* Deep learning detection of Harmful Algal Blooms
* Global Mangrove Monitoring
* Oil Spill Detection
* Offshore Windfarm Monitoring
* Develop ESA product of Particulate Inorganic Carbon
* Biodiversity in the Open Ocean: Mapping, Monitoring and Modeling

### [MLHub and AI/ML in CMR](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_16.00_MLHub.pdf)

Valerie Dixon (NASA) and Michael Morahan (NASA) discussed the MLHub and AI/ML in CMR. The MLHub facilitates an open community common for geospatial training data, machine learning models, and standards to encourage collaboration and share information:

* Python Client allows users to search and download geospatial training data on Radiant MLHub without managing API requests. Users can apply MLHub with other scripting languages using our REST API.
* All Radiant MLHub geospatial training data collections are stored using STAC-compliant catalogs and are exposed through a common API. Radiant Earth is developing the STAC ML Model Extension to the which will empower users to discover and access existing repositories of ML models for various geospatial applications.
* Current content includes 65 datasets and six services.

MLHub and CMR connection: MLHub generates datasets, service models, and metadata. MLHub metadata is placed on the cloud server. IDN scripts pull MLHub metadata for CMR ingest (Associations: MLHub data and service, and CMR data and MLHub data). ML datasets and model data are advertised on NASA AI/ML portal. Users can download data.

Currently in place:

* MLHub: Ready to go.
* Two pathfinder ML records in CMR: A training data record, and model record.
* An AI/ML Earthdata Search Portal: Or you can filter Earthdata Search.
* GCMD Keywords for AI/ML: Discussed at WGISS-54 meeting.

Still to work on:

* Automating ingest of metadata from MLHub to CMR: Coming soon.
* Associating CMR records and ML Training Data as Related Collections to enable cross-discovery.
* ML Model metadata schema – Under Analysis: Pathfinder Model record is a Collection, but prevailing thought is that Training Data are Collections and Models should be Services… or maybe a new schema type?
* Discovery by Service feature in Earthdata Search (and its portals).

Two datasets are registered. There are more within the IDN that could be identified, so they will be adding another 65 soon.

## Federation

### [STAC Best Practice Activity](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/1/2023.04.18_16.25_STAC_Activity.pptx)

Yves Coene (Spacebel for ESA) gave a presentation on the STAC Best Practice activity. He began with a description of STAC. He noted that the multiple options for implementation of STAC lead providers and open-source tools developers to implement arbitrary (different) subsets, causing interoperability issues, raising the question of what WGISS should recommend to be implemented by CEOS data providers?

Most CEOS Best Practices start by identifying scenarios and use cases. If these are not explicit, everyone will have their own interpretation. The group needs to agree on the scope as there are many possibilities, so Yves listed several potential use cases. The community also identifies use cases and topics to address.

The suggested approach is to propose an applicable list of uses cases to narrow the scope of STAC Best Practices. As part of this step a survey was circulated in March to the member agencies. The finding was that a majority (47%) had static and browsable STAC catalog and searchable STAC API catalog. About half have an OpenSearch-based catalogue deployed. Additional questions addressed migration recommendations, and how the BP should be elaborated and published.

The second step is to prepare a draft STAC Best Practices addressing selected use cases by solving issue in one of three ways:

* Recommend or mandate one method over another in the Best Practices
* Raise issue with appropriate STAC standards body to consolidate on single solution
* Create extensions for missing/needed areas (unlikely).

The next step is to obtain SLT feedback and iterate, interacting with relevant for resolving issues (https://www.ogc.org/about/committees/swg/).

In conclusion the STAC ecosystem consists of many “building blocks” supporting many use cases, often in multiple ways, producing an interoperability risk. Implementing all possibilities is unnecessarily expensive and lack of awareness of these possibilities could lead to breakages between systems. STAC clients may not work as intended with STAC services. Proposed solution:

* Adherence to Best Practices will reduce cost and interoperability risks for both client and server implementations.
* Roadmap with three steps; first step is currently underway.
* Preliminary survey results based on 17 answers:

(1) 88% have deployed or will deploy a STAC catalogue

(2) less interest in static catalog (30%), most demanded topics include:

* + - STAC API (70%)
    - Advertise available queryables (70%)
    - Guidance to allow federation (65%)
    - Encode collection and granule metadata, keywords etc. (65%)
    - “Other topics” proposed: e.g., asset-level search, highly recommended extensions, …

(5)(6) Provide migration recommendations OpenSearch to STAC (45%)

(7) Best Practice in GitHub repository

(8) 47% interested to contribute/participate to meetings.

Next steps are to conclude the survey and define “scope” for Best Practice and start preparation of Best Practice.

Doug added that they have an example of interoperability because of the plethora of options STAC gives. OpenSearch is not very popular anymore and maintenance and quality of those APIs is suffering. OpenSearch can thus be replaced with STAC.

Yousuke asked if scope of STAC includes authentication? Doug said that STAC will not provide extensions because it is owned by the user.

Homero asked about a relationship with OGC and Yves replied that STAC relies on OGC specifications.

Makoto noted this will be a key activity for the CEOS Interoperability Framework. Damiano added that this work is still preliminary and it would best to mature the existing implementation of STAC, while keeping communication with the STAC community. Doug stated that the goal is to replace OpenSearch with STAC in CWIC but that would require a BP.

## Outcomes, Actions, Next Steps

Yosuke Ikehata (JAXA) asked the participants to think about adding Data Presentation as a new topic for technology exploration.

Tom asked for clarification on how to differentiate the WGISS BPs and all those findable on the web. Doug replied that the WGISS ones have a narrow focus and should be clearly labeled with the niche they cover, with specific use cases.

Action WGISS-55-16: Doug Newman to brief the CEOS Interoperability Framework team on STAC.

Action WGISS-55-17: Esther Conway to schedule two Jupyter Notebooks workshops (July, September): One on access metadata, and one on technical dependencies.

# Agency and Liaison Reports



## [ASI (Italian Space Agency)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_11.35_ASI%20Agency%20Report.pdf)

A. Mountuori gave an agency report for ASI. He reported that:

* COSMO-SkyMed, CSK-3 deorbiting procedure started on May 2022 and CSG-2 is operational: After successfully passing the test and in-flight qualification stage, the data acquired by the CSG2 satellite are made available to users by the Italian Space Agency and Ministry of Defence. Now, on the same orbit, 3 CSK + 2 CSGCSG mission will be upgraded with third and fourth satellites CSG-3/4 (work in progress). The A new “Open Call for Science” on the use of CSK and CSG data is available on the ASI website:
* PRISMA is operating nominally (https://prisma.asi.it/) and PRISMA SG Phase A activity is on-going.
* Since 20 July 2021, ASI has started the provision of SAOCOM data within the ASI ZoE.

He concluded listing recent journal and conference papers.

## [European Space Agency (ESA)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_1630_ESA%20Agency%20Report.pdf)

Mirko Albani gave an agency report for ESA. He noted an increased budget and its distribution and listed the ESA-developed EO missions, highlighting that ESA manages the Copernicus Space Component, the largest producer of EO data in the world.

Mirko displayed examples of Sentinel products, and discussed the new Copernicus Data Space Ecosystem Service, the next step in the evolution of EO data access. Mirko also discussed the Earth Explorers Development, and the New Earth Explorers Development Earth Observation Heritage Missions at ESA, and concluded with heritage mission data quality improvements, and improvements in data access, along with data volume statistics.

Tom asked if heritage missions go to the cloud? Mirko replied that their mandate is to keep and preserve, so after five years the data are moved to heritage status.

Makoto asked for details on the development plan for scientific missions; Mirko replied that scientists collect requirements and final decisions are taken by a board based on this user group. For Copernicus it is different, since it is under the EC, but decisions are usually based on user requirements.

## [Indian Space Research Organization (ISRO)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_09.00_ISRO%20Agency%20Report.pdf)

T. Sai Kalpana gave an agency report for ISRO. She announced that EOS-06 (Oceansat-3) Ocean Colour and Winds was launched, and described the sensors. Other operational missions are RESOURCESAT-2/2A, EOS-04 and CARTOSAT-2E, EOS-06 , SARAL, and INSAT 3D and 3DR.

Planned missions for the next four years are INSAT-3DS, RISAT-1B, OCEANSAT-3A, RESOURSAT-3 and -3A, NISAR with NASA, ResourceSat Sampler 3S/3SA, and TRISHNA with CNES.

T. Sai Kalpana described the following data resources:

* The Bhoonidhi web portal (EO Data Hub) enables access to archive of Remote Sensing data from 47 satellites, including Indian and foreign remote sensing sensor data acquired over 33 years.
* Bhoonidhi has browse and order (including thematic ordering), VISTA (E06 data visualization), UPGRAM, API services, and Code Lab.
* Bhuvan provides geospatial data and services, including in-situ atmospheric CO2 data and AI/ML value added services.
* VEDAS allows geospatial analysis and AI/ML derived products.
* MOSDAC is the data archival for meteorology and oceanography, with numerous products and services.
* MOSAIC is the MOSDAC platform of services for analysis and interactive computing.
* NICES is the source for multiple terrestrial, ocean, atmospheric, and model derived products.

T. Sai Kalpana confirmed that the new products would be put in the IDN.

## [NOAA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_16.40_NOAA%20Agency%20Report.pptx)

Ken Casey (NOAA) gave an agency report for NOAA. He provided the following summary:

* Update on NESDIS Common Cloud Framework (NCCF)
  + NCCF is a cloud enterprise architecture with a multi-account and multi-Virtual Private Clouds (VPC) deployments continues progress.
  + NESDIS Cloud Archive and Access Team achieved Ver. 1 of the archive services successfully completed end-to-end testing for identified data and is working with the Infrastructure and Solutions Teams to move to production. Initiated scalability testing of archive services. Underlying this new cloud archive is a new, essential component: a knowledge graph (KG).
* Updates from NOAA/NESDIS/NCEI
  + OneStop NOAA Data Catalog V3.0.1 in Production this week, with ~100,000 collections and tens of millions of granules loading (in process).
  + Will replace existing official NOAA Data Catalog once cross-NOAA reviews complete.
  + CoMET updates continue but still no broad public access yet (targeting Oct 2023).
  + NOAA Data Governance is maturing: a new Data Governance Committee finalizing the new NOAA Data Handbook.

Participants thanked Ken for a good explanation of the usefulness and power of knowledge graphs; Ken noted that NOAA is providing use cases.

Makoto mentioned that the US National Science Foundation just released a call for proposals for Proto-OKN Solicitation (NSF 23-571), Building the Prototype Open Knowledge Network (Proto-OKN).

## [Russian Federal Space Agency (ROSCOSMOS)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_11.45_Agency%20Reports%20Roscosmos.pptx)

Andrey Kuklin (ROSCOSMOS) gave an agency report for (ROSCOSMOS). He listed the participation of ROSCOSMOS in CEOS and other international organizations, including the Russian EO systems used in disasters monitoring and participation in the International Charter on Space and Major Disasters.

Andrey described the EO data availability, including search, browse and order through the new geoportal “ROSCOSMOS,” automated data processing and order fulfillment. He also outlined their services and products for forest control, eco-monitoring, agri-monitoring, and industrial services.

## [United Kingdom Space Agency (UKSA)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_16.50_UKSA_AgencyReport.pptx)

Esther Conway (UKSA) gave an agency report for UKSA. She described the goals of UKSA and the nine Catapults, and continued with a discussion of the EO DataHub Pathfinder Project.

In the EO Data Hub project, the National Physical Laboratory (NPL) will lead the development of the data quality catalogue. NPL has led the interaction between the metrology and EO communities, in particular covering cal/val and data quality. The EO Data Hub will be a new ‘single point’ EO data infrastructure that builds on current UK EO assets and brings together the current breadth of UK EO data offerings from public and commercial centres.

The challenge is disparate resources and service offerings, multiple APIs, data formats, license agreements, and operating models. The hub will act as a proxy for the user to mediate access to data and services.

Esther described the Platform development approach and high-level architecture. She noted that an illustration of some of the principles can be found in the past project ESA Digital Twin Earth Climate Explorer.

This new EO DataHub project may have relationships with CEOS Interoperability Framework and be involved with other aspects of CEOS. She suggested interoperability knowledge exchange:

* Pathfinder approach
* Value Public sector, commercial, academic communities
* Project which seeks to balance complexity vs operational sustainability
* Standards-based and open source
* Use Cases for Applications/Platforms drive and inform Hub development priorities
* Build on and extend existing software solutions where advantageous – knowledge exchange
* Maximise use of existing service offerings, avoid re-invention
* Leverage public cloud eco-systems for accelerated development

Makoto mentioned that CEOS wants to develop the New Space area and CEDA has experience with private companies and with relying on industrial research partners. Is this experience something that CEOS will be able to take advantage of?

Makoto welcomed their participation in the CEOS Interoperability Framework. The experience with this new project will provide use cases and examples.

## [USGS Emergency Operations Activities](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/2/2023.04.19_11.35_USGS%20EO.pdf)

Tom Sohre (USGS) outlined the USGS Emergency Operations Activities. In support of Emergency Operations, the Earth Resources Observation and Science (EROS) provides current and accurate geospatial data for local, national, and global hazard support.

Recent USGS support to the International Charter includes 43 new charter events with 219 new data acquisition requests, and liaison activities with a multitude of International Disaster Management entities. Domestic support by USGS includes 31 new non-Charter events with 161 new data acquisition requests, and participation in inter-agency support for hurricanes Fiona and Ian.

Tom described the operational flow for emergency response, and the Hazards Data Distribution System (HDDS). Over the last 12 months, the HDDS systems have distributed 30195 files to the disaster response community. He concluded with a list of the HDDS available platforms.

Andrew Eddy (WGDisasters) commented that this is very useful for their working group to know.

# WGISS Plenary, Part II



## [WGISS Strategic Discussion of the CEOS Interoperability Framework](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_14.00_CEOS%20Interoperability%20Framework%20Discussion.pptx)

Makoto Natsuisaka (JAXA) and Tom Sohre (USGS) led a discussion on WGISS strategy for the CEOS Interoperability Framework. Tom said that he reviewed the CEOS Interoperability Handbook published in 2008, and extracted the following concepts (including 23 recommendations):

* Syntactics
* Semantics
* Levels of Interoperability
* Archive
* Access
* Services
* Technologies

Conversely, the draft CEOS Interoperability Framework contained the following factors:

* Semantic
* Syntactic
* Data Architecture
* Data Accessibility
* Data Quality
* Data Policy
* Overall Coordination

The current proposal contains the following factors:

* Semantic
* Data/Service Technology
* Data Architecture
* Data Accessibility
* Data Quality
* Data Preservation
* Data Policy
* Interoperability IG

|  |  |  |
| --- | --- | --- |
| **2008 Handbook** | **2023 Draft Framework** | **Framework Proposal** |
| Semantics | Semantic | Semantic |
| Syntactic | Syntactic | Data/Service Technology |
| Levels of Interoperability | Data Architecture | Data Architecture |
| Access | Data Accessibility | Data Accessibility |
|  | Data Quality | Data Quality |
| Archive |  | Data Preservation |
|  | Data Policy | Data Policy |
| Services |  |  |
| Technologies |  |  |
| (23) Recommendations | Overall Coordination | Interoperability IG |

Reasoning:

* For ease of management, suggest attempt to align assigned Framework Factors (to WGISS) with WGISS IG’s
* Re-institute the WGISS Interoperability IG (potentially with the Vice-chair as the lead for this).
* Cross-IG coordination, Cross-WG coordination for Interoperability, Use Case development, Communications
* Keep Semantic (because WGCV volunteered to lead; and they, including Peter, have a strong interest in this topic area)
* Rename Syntactic to something like "Data/System Technologies". The term "syntactic" is not referenced within the 2008 document. That "factor" could be adjusted to tie more closely to the system and data technologies and could align with the Technology Exploration IG. “Syntactic interoperability refers to interoperation of the format as well as the data structure used in any stored exchanged information or service between heterogeneous IoT system entities”
* Keep Data Architecture, again because an external group has volunteer to lead and has interest (CEOS ARD OG)
* Keep Data Accessibility and align this to WGISS Data Discovery and Access IG
* Keep Data Quality, again because an external group has volunteer to lead and has interest (WGCV)
* Add Data Preservation and align to WGISS Data Preservation and Stewardship IG
* Keep Data Policy (this may also be a good topic to use for collaboration with SEO)

Discussion:

Regarding the syntactic factor: the original proposal did not seem to have much detail on it besides “data formats,” but it should be kept. Aligning it with an interest group gives some ownership, but the new framework should be technological neutral. The Tech Expo would be a contributor to all.

All agreed that the Data Policy factor should be kept, as national laws support the issue of data delivery, open government data, and transparency; the common issues should be identified, as well as the hurdles. Not a lot of work to be done necessarily, but important to have. Doug added that it has a new dimension, and that is the cloud; there is work to be done there. WGISS does not have technical competence in this area, so suggest that it be led by the SEO. Dave agreed to explore SEO support of this; maybe it could also tie into New Space Task Team.

Decision:

Framework Proposal: Overall effort led by Interoperability IG, led by WGISS Vice-chair.

Factors:

* Semantic
* Syntactic (WGISS)
* Data Architecture
* Data Accessibility (WGISS)
* Data Quality
* Data Preservation (WGISS)
* Data Policy

Support: Nitant, Dave, Damiano, Esther and Mirko verbally support this proposal; all others present agreed.

Next steps:

* Tom to start working on setting up the Data Interoperability and Use Interest Group and start assigning work.
* Develop a draft version roadmap by the end of August, in preparation for the SIT TW. The roadmap could align with the CEOS Workplan timeframe to show progress.
* Begin asking for documentation from each factor lead.
* The framework can evolve, but initial consensus is needed to develop a roadmap. Work that CEOS is already doing should be aligned into the framework.
* Identify the best interoperability use cases for the framework.
* Review the 2008 Interoperability Handbook exists, to build on prior work.

## [Future Meetings](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_15.20_Future%20Meetings.pptx)

Tom Sohre (USGS) discussed the plans for upcoming meetings. He reminded that recent meetings have been:

* WGISS-54: Tokyo, Japan | October, 2022 (joint with WGCV)
* WGISS-55: Córdoba, Argentina | April 18-20, 2023 (joint with WGDisasters)

Planned upcoming meetings are:

* WGISS-56: Paris, France, October 24-26, 2023, hosted by Centre National d’Etudes Spatiales (CNES). Point of contact is Richard Moreno. Information on the meeting will be available on the meeting website <https://ceos.org/meetings/wgiss-56/>
* WGISS-57: To be determined. Planned for March-May, 2024
* WGISS-58: Sioux Falls, USA, October 2024, and joint with WGCV. Hosted by USGS Earth Resources Observation and Science (EROS) Center.

## WGISS Vice Chair Solicitation, and Organizational Matters

Makoto Natsuisaka (JAXA) announced that Nitant Dube (ISRO) has been nominated as the next vice-chair of WGISS. The members present approved the nomination. Nitant shared that he looks forward to continuing to move WGISS forward in the objectives of CEOS.

The Data Interoperability and Use Interest Group will be reinstated. The lead is TBD.

## [Review of WGISS Actions](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_15.45_Summary%20Actions.pptx)

Michelle Piepgrass (JAXA) reported that actions from earlier meetings are closed. The group discussed the actions from the meeting and they were finalized as follows:

**CEOS Interoperability Framework**

* Action WGISS-55-01: Each factor Lead of the CEOS Interoperability Framework to finalize the factor name, coordinator (lead), collaborator(s), description (1-2 paragraphs), scope (included and excluded).
* Action WGISS-55-02: WGISS to work with Thomas Huang to consider the digital twins as a use case in the CEOS Interoperability Framework.
* Action WGISS-55-16: Doug Newman to brief the CEOS Interoperability Framework team on STAC.

**Technology Exploration**

* Action WGISS-55-03: WGDisasters to work with WGISS Technology Exploration Group to incorporate as use cases into the AI/ML white paper the Wildfire Pilot, and perhaps also the Flood (GEO/LEO/SAR) Pilot, Landslide Demonstrator, Volcano Demonstrator, Seismic Hazards Demonstrator.
* Action WGISS-55-04: WGISS to provide WGDisasters updates and/or access to the AI/ML White Paper for their collaboration.
* Action WGISS-55-17: Esther Conway to schedule two Jupyter Notebooks workshops (July, September): One on access metadata, and one on technical dependencies.

**Data Discovery and Access**

* Action WGISS-55-05: WGISS (Yves Coene, Damiano Guerrucci) to invite WGDisasters to provide tools and software developed by WGDisasters so they can be made discoverable to global users through the service discovery.
* Action WGISS-55-06: DSIG/SLT to confirm that content of the [Open-Source Software page](http://ec2-3-208-162-171.compute-1.amazonaws.com/ossi/browse)  on the CEOS website has been correctly ingested in the FedEO operational system that is accessible to end users.
* Action WGISS-55-07: DSIG to ask SEO to remove from the CEOS website the [Open-Source Software page](http://ec2-3-208-162-171.compute-1.amazonaws.com/ossi/browse), upon completion of previous action.
* Action WGISS-55-08: SLT to define a procedure for registration of services/tools into IDN/FedEO (direct registration by users or through circulation of Open-Source Software spreadsheet). SLT to inform WGISS-All on how to register their services and tools.
* Action WGISS-55-09: SLT to examine the emerging needs and possible solutions for user authentication and authorisation in the new scenario for data access. Provide a status report at WGISS-56.

**Data Preservation and Stewardship**

* Action WGISS-55-10: DSIG to prepare a white paper on in-situ data based on the study conducted, and submit it to WGCV for review. Review of the content should be completed by the end of May, and the document should be circulated by end of June if no updates are needed.
* Action WGISS-55-11: DSIG to prepare a white paper on software preservation based on the study conducted. Complete the review of the content by end of May, and circulate by end of June if no updates are needed.
* Action WGISS-55-12: DSIG to circulate the package of WGISS updated White Papers and Reports to WGISS-All to decide if they should be updated or retired by the end of April. The WGISS website should be updated accordingly by the end of May. DSIG should also clearly state the issue date, suggested retirement date, and the type of document (White Paper, technical note, Best Practice). The list should be submitted to the VC leads for their consideration.
* Action WGISS-55-13: Makoto Natsuisaka to ask the SEO and CEO to provide a method of governance on the process for submission and “publication” of CEOS documents (White Papers, Best Practices, Technical Documents, and deliverables from the Work Plan).
* Action WGISS-55-14: DSIG to organize a session on AVHRR at WGISS-56.

**General**

* Action WGISS-55-15: Michelle Piepgrass to find out number of downloads from CEOS website (Best Practice, White Papers, presentations) and provide the results to WGISS Exec.

## [Concluding Discussion](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-55/3/2023.04.20_15.45_Summary%20Actions%20and%20Feedback.pptx)

Tom Sohre led a discussion of the lessons learned from the meeting. All the participants (remote and in-person) were invited to provide feedback. In summary:

Positive feedback:

* **Logistics:** A tremendous thank you to CONAE! The facilities were great, as were the audio/visual, the breaks and hosted events, and the transportation to and from the venue. The agency tour was superb, and very interesting. The meeting was a great opportunity for CONAE to show their activities and facilities, and to exchange information.
* **Joint session:** The joint session was interesting and well planned, and the resulting cross collaboration is very valuable.
* **Content:** It was obvious that the agenda contained expanded time for discussion, leading to good planning for the coming year. The results can be seen already in the actions and future planning.
* **In-person participation:** The meeting provided some with opportunity for reflection based on what was learned. Side conversations were helpful, allowing participants to see where their work is going, and to identify collaborations and sharing of resources. One participant felt that by having presentation each day, it was easier to remain engaged. Others were pleased to have the opportunity to establish old and new friendships and collaborations.
* **Remote participation:** Theremote participants were accommodated well; they remained engaged throughout, which is hard to do for three days. Their dedication made it feel that they were truly part of the meeting.
* **Objectives met:** More discussion time and improved communication to CEOS of WGISS technical expertise.

Areas for improvement:

* **Audio**: Try to solve the technical challenges of the hybrid meeting by exploring audio devices, closed captioning, and taking personal responsibility for being heard and understood.
* **Meeting scheduling**: Consider having one in-person meeting per year, and one or more remote meetings. Consider meeting in person every nine months
* **Content**: Reinstitute the session on announcing newly released datasets and products. In the technology session have a presentation and discussion about the technologies coming up and those that are obsolete. During each session, focus and go in-depth on specific topics, and avoid presentations less than 10 minutes, or cutting presentations short.

## Meeting Conclusion

Makoto Natsuisaka thanked CONAE for their hosting of the meeting. All the arrangements were excellent and well-received. He also thanked the in-person and remote participants for their active collaboration, and wished everyone safe travels home.

# Glossary of Acronyms

AI Artificial Intelligence

API Application Programming Interface

ARD Analysis Ready Data

AWS Amazon Web Services

CC Climate Change

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

CMR Common Metadata Repository

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DC data cube

DIF Directory Interchange Format

DLT Digital Linear Tape

DOI Digital Object Identifier

ECV Essential Climate Variable

EO Earth Observation

GCMD Global Change Master Directory

GEO Group on Earth Observations

GEOSS Global Earth Observation System of Systems

GFOI Global Forest Observations Initiative

GIS Geospatial Information System

HPC High Performance Computing

IDN International Directory Network

ISO International Standards Organization

LEO Low Earth Orbit

LSI Land Surface Imaging

ML Machine Learning

NRT Near real-time

OGC Open Geospatial Consortium

PI Persistent Identifier

POC Point of Contac

RS Remote Sensing

SEO Systems Engineering Office (CEOS)

SDCG Space Data Coordination Group

SIT Strategic Implementation Team

SLT System Level Team

SWG Standards Working Group.

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