MINUTES

OF THE

51st MEETING

OF THE

CEOS WORKING GROUP ON   
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Held virtually on

April 20-22, 2021

Table of Contents

[1 WGISS Plenary Session, Part I 5](#_Toc75715101)

[1.1. Introduction 5](#_Toc75715102)

[1.2. WGISS Chair Report 5](#_Toc75715103)

[1.3. CEOS Executive Officer (CEO) Report 5](#_Toc75715104)

[1.4. Systems Engineering Office (SEO) Report 5](#_Toc75715105)

[1.5. GEO-SEC Report 6](#_Toc75715106)

[2 Data DISCOVERY and ACCESS 7](#_Toc75715107)

[2.1. Introduction 7](#_Toc75715110)

[2.2. General Reports 7](#_Toc75715111)

[2.2.1. International Directory Network (IDN) 7](#_Toc75715112)

[2.2.2. CWIC and CWIC Evolution 7](#_Toc75715113)

[2.2.3. FedEO 7](#_Toc75715114)

[2.3. OpenSearch Best Practices Document (Recent Updates) 8](#_Toc75715115)

[2.4. QC-MMS – Experience with Quality Metadata, Catalog and Jupyter Notebook 8](#_Toc75715116)

[2.5. NASA EOSDIS STAC 8](#_Toc75715117)

[2.5.1. State of STAC 8](#_Toc75715118)

[2.5.2. STAC and Harmony 8](#_Toc75715119)

[2.5.3. STAC’s Potential for Federated Search 9](#_Toc75715120)

[2.6. USGS Migration to Cloud 9](#_Toc75715121)

[2.7. NASA Cloud Data Access 9](#_Toc75715122)

[2.8. ESA Extended Data Access 10](#_Toc75715123)

[2.9. ESA Earth Observation Identity Access Management 10](#_Toc75715124)

[3 Data INTEROPERABILITY and USE 11](#_Toc75715125)

[3.1. Introduction: Cloud Formats Best Practice or White Paper Development 11](#_Toc75715128)

[3.2. Technical Readiness of ARD 11](#_Toc75715129)

[3.3. Analysis Ready Satellite Data 11](#_Toc75715130)

[3.4. SEO Perspective on CEOS Satellite Data in the Cloud 11](#_Toc75715131)

[3.5. ESA Digital Twin Earth Climate Explorer Project on JASMIN 12](#_Toc75715132)

[3.6. Cloud-Native Earth Observation Data and Metadata Standards 12](#_Toc75715133)

[3.7. CEOS EAIL Update 12](#_Toc75715134)

[3.8. Zarr Cloud Format in Open Data Cube for Hyperspectral Data 13](#_Toc75715135)

[3.9. A Parquet Cube Alternative to Store Gridded Data 13](#_Toc75715136)

[3.10. Data Interoperability and Use Discussion, Summary of Actions 14](#_Toc75715137)

[4 TECHNOLOGY EXPLORATION 15](#_Toc75715138)

[4.1. Introduction 15](#_Toc75715141)

[4.2. WGCapD Webinar Toolkit 15](#_Toc75715142)

[4.3. Jupyter Notebook for WGCapD Webinar 15](#_Toc75715143)

[4.4. CSIRO Jupyter Notebook Experience video link 15](#_Toc75715144)

[4.5. Developing a Jupyter Notebook Best Practice 15](#_Toc75715145)

[4.6. Future Topics for Technology Exploration 15](#_Toc75715146)

[5 Data PRESERVATION and STEWARDSHIP 16](#_Toc75715147)

[5.1. CAL/VAL Maturity Matrix 16](#_Toc75715150)

[5.2. NOAA Maturity Matrix Self-Assessment Tool 16](#_Toc75715151)

[5.3. Persistent Identifiers 16](#_Toc75715152)

[5.3.1. PID Open Points Discussion 16](#_Toc75715153)

[5.3.2. NASA PID 17](#_Toc75715154)

[5.4. Archive Holdings and Technology Session 17](#_Toc75715155)

[5.4.1. ESA 17](#_Toc75715156)

[5.4.2. NOAA 17](#_Toc75715157)

[5.4.3. STFC/CEDA 17](#_Toc75715158)

[5.4.4. USGS 17](#_Toc75715159)

[5.4.5. NASA 17](#_Toc75715160)

[5.4.6. JAXA 18](#_Toc75715161)

[5.4.7. ISRO 18](#_Toc75715162)

[5.5. International Cooperation on AVHRR Data 18](#_Toc75715163)

[6 Agency and Liaison Reports 19](#_Toc75715164)

[6.1. 19](#_Toc75715167)

[6.2. NASA 19](#_Toc75715168)

[6.3. NOAA 19](#_Toc75715169)

[6.4. CSA-CCMEO 19](#_Toc75715170)

[6.5. USGS 19](#_Toc75715171)

[6.6. CNES 19](#_Toc75715172)

[7 WGISS Plenary, part II 20](#_Toc75715173)

[7.1. Future Meetings 20](#_Toc75715175)

[7.2. WGISS Summary and Discussion 20](#_Toc75715176)

[8 Glossary of Acronyms 22](#_Toc75715177)

ASI Francesco Tataranni

CAS/RADI Qianqian Cao

CEOS Executive Officer Marie-Claire Greening

CNES Richard Moreno, Jean-Michel Zigna

CONAE Homero Lozza, Josefina Ortero

CSIRO Robert (Rob) Woodcock (WGISS Chair), Matt Paget, Michelle Piepgrass (WGISS Secretary), Peter Want

DLR Jonas Eberle, Thomas Ochs, Charlotte When

EC Peter Strobl

ESA Mirko Albani (WGISS Chair), Pavan Kumar Alikana (Rhea Group), Veronique Amans, Paolo Castracane, Yves Coene, Andrea Della Vecchia, Philippe Goryl, Damiano Guerrucci, Daniele Iozzino (Rhea Group), Marco Leonardi, Iolanda Maggio (Rhea Group), Philippe Mougnaud, Sabrina Pinori, Giuseppe Troina

GA Alex Leith

GEO Paola de Salvo, Gilberto Camara, Douglas Cripe

ISRO Nitant Dube, Ravi Kant, Ankitha E., Sai Kalpana

JAXA Makoto Natsuisaka, Yousuke Ikehata

NASA Andrew (Andy) Mitchell, David Borges, Lauren Childs-Gleason, Diane Davies, Liping Di, David Green, Matt Hanson, Kim (Holloway), Brian Killough (CEOS-SEO), Dawn Lowe, Christopher (Chris) Lynnes, Michael Morahan, Douglas Newman, Hampapuram Ramapriyan (Rama), Kenton Ross, Nancy Searby, Amy Steiker, Min Wong.

NOAA Kenneth (Ken) Casey, Prasanjit Dash, Paul DiGiacomo, Merrie Neely, Diana Ostrenga, Nancy Richey, Martin Yapur,

NOAA & NASA Aijun Chen

SANSA Edwin Magidimisha

UKSA Esther Conway, Steve Donegan, Ag Stephens, Phil Kershaw, Richard Smith

USGS Kristi Kline, Steve Labahn, Ryan Longhenry

# WGISS Plenary Session, Part I

## Introduction

Robert (Rob) Woodcock (CSIRO), WGISS-Chair opened the WGISS-51 meeting, thanking the participants for their attendance to the virtual meeting. Rob gave the highlights of the agenda, and explained that each presenter would have five minutes to present a summary of the report, followed by a short comment/question period. He noted that this is a working meeting, and asked everyone to participate verbally or using the chat feature.

## [WGISS Chair Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_Welcome.pptx)

Rob began his report with a call for nomination for Vice Chair of WGISS for the term 2021-2023 (followed by the position of Chair (2023-2025). Rob listed the following WGISS activity highlights.

Discovery and Access: OpenSearch Best Practices (update), STAC, Cloud data access.

Data Interoperability and Use: Cloud formats and data use, CEOS Earth Analytics Interoperability Lab (update), and CEOS Interoperability Terminology (endorsed).

Technology Exploration: Jupyter Notebooks for EO Webinars and Best Practice

Data Preservation and Stewardship: Persistent Identifier (PID) Best Practice, and Maturity matrix – self-assessment tool, cal/val.

CEOS Future Data Architecture (FDA): “Now” Data Architectures (USGS Collection 2 on AWS, and others). WGISS will be updating action, augmenting best practices of discovery systems and cloud data formats: COGS, Zarr.

CEOS Analysis Ready Data (ARD): Card4L activities include surface reflectance and temperature, aquatic reflectance and night-time light surface radiance. There has been a lot of quite foundational work, improving the metadata that is available to describe the data, useability across platforms.

The goal of WGISS is to make satellite data more useable and accessible. WGISS welcomes engagement from all CEOS members.

## [CEOS Executive Officer (CEO) Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_CEOReport.pptx)

Marie-Claire Greening, CEOS Executive Officer, stated that her role is to work with all parts of the CEOS family, with an external and internal focus. Cooperation between member agencies and organizations is essential, and contributions are valued on a best effort basis. She described the aspects of the CEOS Chair Theme (Space-based Earth Observation Data for Open Science and Decision Support) priorities that relate to the work of WGISS:

Data Deployments in the Cloud – Increase the number of free/open datasets in public computing clouds for improved access and use.

Earth Analytics Interoperability Lab (EAIL) – Develop the first release of the EAIL to benefit several CEOS initiatives (e.g., COAST, Disasters, Asia-RiCE) by allowing users to combine multiple interoperable datasets from the land and ocean domains in an open and shared environment.

Open Data Cube Sandbox – Release the first ODC Sandbox as a contribution to the Open Earth Alliance GEO Community Activity and the GEO Knowledge Hub to demonstrate global CEOS data access via computing clouds using a free/open programming framework. (FDA-17-02)

The CEOS Work Plan 2021-23 allows updates from the tracking tool which is easy to use. Updates are encouraged and welcome.

## [Systems Engineering Office (SEO) Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_SEO_Report.pptx)

Brian Killough (NASA, SEO) discussed highlights of the Systems Engineering Office (SEO) activities:

Google Colab Sandbox: The SEO recently completed the first beta (internal) release of an Open Data Cube sandbox that runs on Google Colab and connects to Earth Engine datasets. This free/open tool can be used for Python notebook training as it can scale to unlimited users. This does require a google account and an earth engine authentication to prevent compromising the system. It is scalable.

Earth Analytics Interoperability Lab: The SEO is working with Rob Woodcock to support data and algorithm testing for CEOS initiatives (COAST, Disasters, Rice Monitoring, GHG, DEM evaluation).

Regional Data Cubes: The SEO is working closely with stakeholders in the Americas (ECLAC, AmeriGEO) and the Pacific Islands (SPC) to pursue regional data cubes similar to what was accomplished in Africa. A survey at the Americas meeting concluded that this type of initiative would help them. Both of these will be quite different than Africa.

Open Earth Alliance (OEA): A new non-profit entity linked to GEO (community activity) that can expand the impact of the Open Data Cube and other data technologies (https://openearthalliance.org). The OEA recently released a new user forum to support the growing list of ODC users (<https://www.openearthalliance.org/forum>). The intent is to gather funding to expand the impact of the Open DC.

AWS Cloud ... Landsat Collection-2 is now on AWS-West along with Sentinel-2 (processed by Element-84). Both analysis-ready datasets take advantage of Cloud-Optimized GeoTIFF (COG) data formats and Spatio-Temporal Asset Catalog (STAC) metadata formats. This will support a number of global data cube projects and provide an efficient cloud-based solution.

Sentinel-1 Data Flows ... The SEO worked with DE-Africa to co-fund (via Sinergise) the development of a Sentinel-1 Data Cube processing pipeline that complies with ARD specifications and uses COG format and STAC metadata. This is only for backscatter intensity data cubes. This tool allows the generation of global S1 data cubes anywhere in the world. Contact the SEO if you want to create and test S1 cubes using this tool. <https://apps.sentinel-hub.com/s1-card4l/>

Brian noted that in situations of diverse terrain and lack of knowledge of terrain, drones are now being used.

## GEO-SEC Report

Gilberto Camara (GEOSEC) gave the report from the point of view of the GEO-SEC. He noted that the goal that is of interest to WGISS is to support end-to-end processing, where the ARD is obtained and placed in a cloud collection, from which a data cube can built. There is a conceptual definition of EO DCs that is entirely defined by a spatio-temporal extent, a vector of attributes with domain and co-domain and a field function. Conceptually the user should be able to touch any pixel and get a time series, or identify a point in time and obtain a x-y location.

All of this encompasses a huge set of operations, where the complexities are hidden from the user, without a proliferation of code that does similar operations, following a (yet to be determined) standard.

Alex Leith commented that Open EO is a conceptual work flow that is transformed to the API, but has to work within the Google file system. Adding a standard API to the ODC will only make it stronger.

Jonas Eberle noted the OGC testbed program, which the task for implementing DC API prototype.

# Data DISCOVERY and ACCESS



## Introduction

Damiano Guerrucci (ESA) introduced the Discovery and Access session.

## General Reports

### [International Directory Network (IDN)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_IDN_Report.pptx)

Michael Morahan (NASA) gave the International Directory Network (IDN) report. He presented CEOS/IDN collaborations with NOAA, ESA, and ISRO, and listed recent IDN Metrics.

Michael noted that NASA has begun to add a Data Usage Policy statement to all collection records. All providers are encouraged to add Data Usage Policy information, as a link or text, to their datasets in the CMR for the IDN.

Michael also discussed recent updates to the GCMD Keyword Viewer.

### [CWIC and CWIC Evolution](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_CWICReport_Evolution.pptx)

Minnie Wong (NASA) discussed CWIC and the CWIC evolution. She summarized the current CWIC implementation as follows:

In mid-April 2021, CWIC functionality transitioned to NASA’s Common Metadata Repository (CMR) and the server at USGS was decommissioned.

The new CWIC architecture provides a mechanism to achieve 'federated discovery' using a centralized collection discovery API and a number of federated granule discovery APIs.

Each CWIC provider has a granule discovery API and the linkage between CMR collection results and the CWIC provider granule APIs is direct (rather than through the CWIC server mediator in the past).

CEOS Best Practices OpenSearch Descriptor Documents (OSDD) were moved from CWIC server to CMR for data partners that did not adhere to CEOS Best Practices for OSDDs. Partners that have working OSDDs are included into CWIC via tags.

The ultimate goal is for ALL CWIC data partners to manage their own OSDDs and are included into CWIC via tags.

CSW Standard support discontinued due to low usage.

New partners and new data from existing providers will continue to be onboarded via the International Directory Network (IDN) which is built on CMR.

CWIC Provider Holdings are available at <https://access.earthdata.nasa.gov/holdings>. Inventory metrics coming soon.

CWIC data discoverable via CMR OpenSearch API – using centralized collection search and federated granule search according to the CEOS OpenSearch Best Practices.

The CWIC Search Portal offers the same user search and discovery look-and-feel as NASA’s Earthdata Search. As the CEOS branded CWIC Portal, it provides collection discovery and granule search for CWIC data partners’ data, <https://search.earthdata.nasa.gov/portal/cwic/search>.

Rob noticed that the portal has a button about AWS cloud. Doug replied that the CMR data property identifies if available on the cloud.

Ken noted that the NOAA OneStop UI is similar - it looks through the access URLs and can highlight different kinds of access URLs with different icons and interface features.

### [FedEO](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_FedEO_Report.pptx)

Yves Coene (Spacebel, ESA) and Andrea Della Vecchia (Randstad, ESA) gave a presentation on the status of FedEO. They discussed ongoing integration activities and gave current metrics and DIF-10 export status, having 30% more granules access. They are currently ingesting ISO data records.

The STAC interface evolution since WGISS-50 includes drill-down capability in STAC catalogues via instruments and earth topics, per year/month/day. They are also able to drill down according to SKOS concept hierarchy (e.g., instruments) to find collections.

The FedEO evolution activities (EOVOC) extends the previous WGISS work. The Draft Engineering is in progress, collecting feedback, comments, and prototype implementation feedback. Next steps are to take into account CEOS WGISS SLT partners’ feedback, completion and refinement of models, and parallel implementation by adding RDF HTML and JSON-LD response types to FedEO catalogue interfaces.

The SKOS hierarchy is based on an existing nomenclature: GCMD with some modifications. The modifications can be made available for inclusion in the GCMD.

## [OpenSearch Best Practices Document (Recent Updates)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_CEOSOSBPupdate.pptx)

Minnie Wong (NASA) discussed the status and recent updates of the OpenSearch Best Practices Document. It was updated to version 1.3 on 12 November 2020 and approved in March 2021; available on the CEOS WGISS website. Updates in the document include a new recommended requirement and one recommended attribute:

CEOS-BP-002D is a new recommended requirement for specifying a maximum temporal range in the OSDD.

CEOS-BP-003 adds a new recommended attribute regarding service level search query URL to the existing collection and granule level search query URL.

Further suggested updates to the document include:

Should the maximum temporal range be a required requirement?

Should spatial/bounding box be a requirement or a recommendation?

## [QC-MMS – Experience with Quality Metadata, Catalog and Jupyter Notebook](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_QCMMS_Experience.pptx)

Yves Coene (Spacebel, ESA) and Andrea Della Vecchia (Randstad, ESA) discussed the QC-MMS Project:  
Experience with Quality Metadata, Catalog and Jupyter Notebooks. Yves described the overall approach and described the QI Catalogue; its purpose, its functional view, its interfaces, its product metadata, and its quality metadata. Yves also discussed the QA Report Manager (Jupyter Notebooks).

Yves concluded that the concept demonstrated with QI Catalogue containing metadata inserted by the MapRadix QC Manager LMS processing chains for HRL Imperviousness land product and HRL Forest Change land product.

Generic standard-based mechanism extending standards-based metadata and catalogue environment.

Quality indicators (QI) based on W3C DQV representation are domain specific but can be replaced by other QI using same representations.

Lessons learned:

Steep learning curve to interact with QI Catalogue from Python-based processor software (QC Manager), adopted OGC metadata formats etc.

Learning curve to interface from Jupyter notebook (report template) to Catalogue.

JSON Schema-based metadata validation much less precise than XML Schema-based validation. JSON payload validations too limited (extensible by default) to avoid integration issues.

Yves listed plans for future work.

Rob commented that the plans for future work are a great opportunity to bring such indicators into the ARD strategy update over the coming months.

## [NASA EOSDIS STAC](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_NASA_STAC.pptx)

### [State of STAC](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_NASA_State%20of%20STAC.mov)

Matt Hanson (Element 84, NASA) discussed CMR-STAC and STAC specification, the CMR-STAC catalog, browser and API. He noted that the final version would be released at end of month and confirmed that STAC is a proxy in front of CMR, and that a new STAC library is open.

### [STAC and Harmony](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_NASA_STAC%20and%20Harmony.mov)

Amy Steiker (NSIDC, NASA) and Doug Newman (Raytheon, NASA) discussed the Harmony STAC catalog. The opportunities made possible by cloud-accessible data include:

* Consistent access patterns to holdings
* Data reduction services allow users to request only the data they want, in the format and projection they want.
* ARD and cloud access will help reduce time-to-science.
* Community development helps reduce the barriers for re-use of code and sharing of domain knowledge.

The STAC catalog is generated for each Harmony asynchronous request. STAC items contain metadata reflecting transformed output extents. Direct access to Harmony outputs utilizing STAC libraries.

### [STAC’s Potential for Federated Search](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_NASA_STAC%20Potential%20Federated%20Search.mp4)

Doug Newman (Raytheon, NASA) discussed the potential of STAC for federated search. Doug remarked that embedding links to distributed catalog APIs in the CMR STAC API should be really simple to do. The advantages are that federated search becomes ‘cloud-friendly’, with simpler coordination. However, it is less flexible to develop, there is no OSDD concept, best practices and guidelines need to be established, and risk of STAC-adherence of federated APIs (similar risk to Open Search).

Richard Smith (UKSA) noted that they have spoken to Radiant Earth about the facet discovery issue and they are open to development in the API to handle that. It should be possible to extend the query spec to report on facet values and counts, in a similar way to OpenSearch. The lack of free-text search in STAC is a concern.

Yves commented that FedEO (ESE-ERGO) allows the STAC API and OpenSearch API to use the facet type parameters from the SRU extension of OpenSearch; some information of how to do that in GeoJSON responses (such as STAC) are the public report OGC 19-020r1.

Matt Hanson (Element 84, NASA) added that the goal is to add a capability for collection search using STAC indexes, crawl APIs in a federated search.

Chris Lynnes (NASA) suggested raising this topic to the OGC community.

## [USGS Migration to Cloud](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_USGS_MigrationToCloud.pptx)

Kristi Kline (USGS) discussed USGS Landsat migration to the cloud. The project scope is:

Modernize Processing, Access, and Distribution of Landsat Data

* + Change from a primary business model of downloads to enabling access to the full archive
  + Enable users to interact with the data in an integrated environment
  + Ensure provenance and data stewardship

Key Project Objectives:

* + Establish an enterprise cloud environment for Landsat
  + Enable access to Collection 1 Level-1 and Level-2 in the cloud
    - Replicate Collection 1 Level-1 and establish operational data management procedures
    - Demonstrate global scale production of Landsat data in the cloud through production of Level-2 products using a cloud framework
    - Process Landsat archive in 1-2 months rather than 9-12 months
  + Establish modern access and visualization tools to access data
  + Establish an Environment and System to Produce and Enable Landsat Collection 2 in the cloud
  + Demonstrate key science use cases exploiting Landsat data

Kristi noted that the tool v 2.0 will be demonstrated to the USGS working group; WGISS will be invited.

## [NASA Cloud Data Access](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_NASA_Cloud_Data_Access.pptx)

Chris Lynnes (NASA) discussed NASA cloud data access. He explained that:

Hosting data in the Cloud enables users to analyze the data in place, with no data movement or management required.

Hosting the data in one virtual place facilitates multi-dataset studies and scientist collaboration.

NASA is developing a framework to support both analysis-in-place and traditional data download access to data.

Chris described the Multi-Mission Algorithm and Analysis Platform (MAA)

Prototype innovative data system to enable open science in cloud. It was implemented initially for Biomass: Lidar, Synthetic Aperture Radar, but is adaptable to other disciplines.

Provides seamless access to satellite, airborne and field data from European Space Agency and NASA

Scale user algorithms from small regions to global scale

Enable collaboration on calibration and validation of higher-level science products

Chris noted that users could benefit from some assistance in moving their analyses to the cloud.

## [ESA Extended Data Access](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_ESA_Extended_Data_Access.pptx)

G. Troina (ESA) and Andrea Della Vecchia (Randstad, ESA) discussed ESA extended data access. They discussed the

* Data Portfolio
  + ESA Heritage Missions
  + ESA Earth Explorers
  + Third Party Missions
* Description of the architecture
  + Layered architecture
  + EO-CAT and FedEO
  + PDGS DataCube
* Evolution
  + ESE-ERGO
  + 3D interfaces (3dFlus)
  + VRE (Aeolus)

In conclusion, this is modular, layered, common architecture ready to fulfil requirements of new missions (e.g., EarthCARE, Biomass, Flex…) with standardization and interoperability. There is a common catalogue in order to increase visibility throughout discovery services. The evolution towards improved data visualisation and processing includes ESE/ERGO, 3D over time series, VRE, and AI.

## [ESA Earth Observation Identity Access Management](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/1.Tuesday/2021.04.20_ESA_EOIdentityAccessManagement.pptx)

Damiano Guerrucci (ESA) and M. Leonardi (Rhea, ESA) discussed ESA’s EO identity access management (EOIAM), which is the new ESA Earth Observation Identity and Access Management system:

Improving the user experience when accessing ESA EO data and related services

Compliant with GDPR

Based on a core an open-source product, the WSO2 Identity Server

Allowing implementation of federations

Offering strong potential for extensions and interoperability

EOIAM benefits include:

ESA/Maintainer side

* + EO IAM System has at its core an open-source product, the WSO2 Identity Server, on top of which a separate plugin component has been developed to comply with ESA’s needs
  + Provides all the authentication functionalities towards ESA Service Providers along with user management capabilities
  + Support to federations and multi-protocol (SAML 2.0, OAuth, OIDC)
  + Configuration changes (user and SP) moved to operational teams
  + Has two user interfaces, an end user interface and operator web interface having a dedicated administrator console
  + Strong potential for extensions and interoperability

ESA/Common Services side

* + Enables M2M authentication (e.g., DataCube, WCS)
  + Out of the box compliancy with Data Protection Policy
  + Look & feel harmonized with ESA TellUS and Earth Online
  + No changes necessary done on the services for transitioning to EO IAM

# Data INTEROPERABILITY and USE



## Introduction: Cloud Formats Best Practice or White Paper Development

Robert Woodcock (CSIRO) introduced the session, noting that an expected outcome is a Cloud Formats Best Practice or White Paper Development.

## [Technical Readiness of ARD](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_Technical%20readiness%20for%20CARD4L.pptx)

P. Goryl (ESA. WGCV Chair) discussed the technical readiness for CEOS Analysis-Ready Data for Land (CARD4L). He noted that ARD represents a growing trend but definition varies greatly across user groups and data providers. The CEOS Land Surface Imaging Virtual Constellation (LSI-VC) was tasked in 2015 to “Define intercomparable ARD products within the context of land surface imaging” 🡪 CARD4L.

CEOS ARD focuses on interoperability, common data formats, inter-calibration, validation and inter-comparison strategies. CARD4L are base for CEOS Data Cube (CDC) architecture:

time-series multi-dimensional stack of spatially aligned pixels

improves computation time vs typical scene-based approaches

The ARD evolution is beyond land: Atmospheric subgroup, Aquatic, Nightlight Radiance, Lidar Terrain and Canopy Height, SAR Geocoded Single-Look Complex (SLC), Interferometric Radar.

CARD4L at ESA:

Surface Reflectance: Sentinel-2 compliant (with recent GRI)

Surface Temperature: ATSR almost compliant; Sentinel-3/SLSTR

Normalized Radar Backscatter (ortho-rectification); Sentinel-1/(A)SAR – work in progress using new Copernicus DEM

[Philippe conclude saying that t](res://\\\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20118%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%207733260%3c/onRightClick%3e%22%3e%3c/A%3e)he Sentinel-2 CARD4L compliant will be available (open and free) through the Copernicus Data Hub.

Chris Lynnes raised the question of how to put Lidar point clouds in a data cube, with its combination of high spatial resolution but very sparse sampling across orbits. [Nitant Dube](res://\\\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20108%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%207077900%3c/onRightClick%3e%22%3e%3c/A%3e) (ISRO) noted that possible options could be as POINT cloud of reflectivity.

## [Analysis Ready Satellite Data](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_Analysis-Ready-Satellite-Data.pdf)

Chris Lynnes (NASA) gave a presentation on Analysis Ready Satellite Data. He discussed level 1, 2, and 3 data, and how to make it analysis-ready. He noted that the considerations for Level 3 data are spatial resolution degradation, suppression of extreme values, systematic bias over time in heterogeneous pixels, day boundary artifacts, difficulty assigning numerical uncertainty and quality to pixels. There is also irretrievable information loss when reprojecting from the original L3 projections.

Phillipe remarked that there is a "contradiction" between L1 and ARD: L1almost by definition is not ready for analysis (you need to do processing before using them - atmospheric correction, or other correction).

## [SEO Perspective on CEOS Satellite Data in the Cloud](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_CloudFormats_SEO.pptx)

Brian Killough (NASA, CEOS SEO) discussed SEO Perspective on CEOS Satellite Data in the Cloud.

There are many TRENDS in cloud-based satellite data: growing data volumes in the cloud, ARD formats, Jupyter notebooks, many vendor options (Amazon-AWS, Google-GCP, Microsoft-Azure), increase use of radar data, open science.

STAC and COG formats are becoming more common; STAC is great for ODC data indexing and COG formats are great for loading subsets of large data files.

Suggestions for the future:

* + Discourage downloading data and use data in the cloud
  + Encourage the use of the cloud computing through CEOS prototypes (e.g., EAIL, ODC-Sandbox)
  + Promote the production and use of ARD with STAC and COG formats
  + Participate in Python notebook training events and focus on “simplicity”
  + Demonstrate the value of radar data by working on ARD processing, data access, and training.

Jonas Eberle commented that it is important to standardize the properties and assets of STAC for specific missions (data collections). As an example, Element84’s Earth search API provides "sentinel:product\_id" and Microsoft STAC API provides "s2:product\_uri" as property with the same content. In terms of platform interoperability, it would be good to standardize these property names to serve a dedicated and fixed set of properties for each of the STAC items (e.g., developing specific STAC extensions for missions).

Nitant added that, in order to access multiple clouds, there is an authentication step for each one. Chris noted that a number of groups are working on this, including cloud computing foundation.

Rob noted that the SKOS linkage is an interesting option to standardize the vocabularies across CEOS.

Andy Mitchell (NASA) raised the issue of “data sprawl” within the same cloud; several copies of the same data in a given cloud.

Philip Kershaw (UKSA) added that the ESA EO Exploitation Platform Common Architecture is another thing to look at for cloud interoperability.

## [ESA Digital Twin Earth Climate Explorer Project on JASMIN](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_ESA_DTEP-JASMIN.pptx)

Phil Kershaw (UKSA) and Ag Stephens (NCEO/UKSA) discussed ESA Digital Twin Earth Climate Explorer Project on JASMIN. These are the highlights:

DTEP Climate Explorer: JULES Land-Surface model - assimilation of EO data with ML emulation – enable real-time use by decision-makers.

JASMIN platform: Data-intensive computing + community cloud hybrid

Develop a Strategy for CMIP data in Object Store for JASMIN

Investigating functionality/performance of object store formats: NetCDF – using "range-gets" // Zarr // S3-netCDF

Other projects using JASMIN Object Store with Zarr: ESA CCI Knowledge Exchange

Object Store issues: CMIP6/PANGEO/Zarr, Data Preservation, Search and cataloguing (Intake/STAC etc), Heterogeneous storage (tape/POSIX/object store)

Phil clarified that, compared to a data cube, Digital Twin is a much larger European program involving ESA, EUMETSAT and ECMWF. Data Cubes could be a part of it but it has a much larger remit. Nitant asked if EO data is assimilated in the model forecast and used in Digital Twin? Philip replied that Digital Twin Earth has a broad remit including model, observations and re-analysis data.

Kenneth Casey (NOAA) stressed that NOAA’s work was done in the context of their operational algorithms, which are largely in C/C++/Fortran where Zarr support is lacking compared to the Python support. Their report does highlight some of the Zarr advantages, for example when doing data-oriented research in Jupyter notebooks or when developing new algorithms, where you can take advantage of things like Zarr's support for multi-threaded applications. Some of the challenges using Zarr might be more particular to their environment, where there are hundreds of existing operational algorithms to deal with.

## [Cloud-Native Earth Observation Data and Metadata Standards](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_DEA%20Data%20Standards.pptx)

Alex Leith (GA) gave a report on cloud-native EO data and metadata context, in the context of the Digital Earth Australia and Digital Earth Africa projects. His key points are:

* Cloud Optimised GeoTIFFs remove the need to pre-tile data for analysis
* Spatio-Temporal Asset Catalog metadata provides a standard to index from
* Indexing STAC into the ODC means that on-demand global data cubes are easy
* Broad area analyses are accessible to more people when data is on the cloud.

## [CEOS EAIL Update](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_EAIL.pptx)

Robert Woodcock (CSIRO) gave update on the CEOS Earth Analytics Interoperability Lab.

Engagement Status:

15 registered users across four CEOS projects. Expecting to double the number of users. Not meant for large number of projects. The projects are Disasters, COAST, DEMIX Asia-RiCE, and OEA community forum for EAIL (open earth alliance is providing support).

Interoperability Status:

(Amazon) US East sample region data: USGS Collection 2 – used-in-place on AWS, Element 84 Sentinel 2 COGS – used-in-place on AWS, Sinergise S1 ARD Service – ordered-and-stored, fee for service, and all confirmed or under assessment for CARD4L

EASI Training notebooks EAIL tailored

Dask Gateway for scalable computing (Larger nodes, GPU – customisable)

Strategy and next steps

Some progress since WGISS 50 – in particular engagement TO working groups (get from detail)

Proposed interop BP development:

* + Cloud access/use for CEOS – metadata and services
  + Cloud formats BP
  + Cloud data lifecycle BP

EAIL and Jupyter Notebook BP

## [Zarr Cloud Format in Open Data Cube for Hyperspectral Data](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_Zarr_Cloud_Format_in_ODC_for_hyperspectral_data.pptx)

Peter Wang (CSIRO) gave a presentation on Zarr cloud format in ODC for hyperspectral data. The project goal is to:

Develop an EO data analytics platform.

Provides access to various types of EO data, including hyperspectral and lidar.

For a range of internal projects.

The project experience is that it supports storage of high dimensional data and storage of hierarchical/nested data. Dask with threads works well with Zarr in increasing IO performance, and similar performance to COGs for 2D. It is suspected that reading 3D Zarr as one 3D read gives better performance compared to COGs.

Lessons Learned:

Zarr creates too many objects, especially if chunk size is small.

Optimal compressed chunk size is approximately 5-20 MB. This results in reads being optimized for cloud storage.

Consolidate the metadata in a single object. This prevents IO overhead in reading the distributed metadata information for a single Zarr.

Rasterised point-cloud data compresses well, but too much overhead.

Representing point-cloud data natively should work well e.g., recursive hierarchy of arrays within a Zarr.

Possible improvements

Zarr chunk handling: Option to store chunks in a single object or multiple objects.

Chunk storage in more detail.

* + Option 1: Contiguous storage of compressed chunks.
  + Option 2: The hybrid - Binned Contiguous storage of compressed chunks.

Decouple metadata from chunk storage.

To do an efficient geospatial subset of a point cloud in Zarr, R-tree style indexing with a space filling curve. R-tree style to efficient index the sparse data. Space filling curve like hilbert to map nD to 1D so points in nD space is stored close in 1D space. This plus a recursive structure of points indexed by a R-tree style index might work. Can have sparse + dense data in the same index.

## [A Parquet Cube Alternative to Store Gridded Data](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_CNES%20Parquet%20Cube%20CLS.pdf)

Jean-Michel Zigna (CLS, CNES) presented on a parquet cube alternative for store gridded data. He stated that this Parquet Cube Alternative is good candidate to face the data analytics and modeling in cloud environment for gridded data. To get good performances in storage and processing:

Parquet storage size is around half of the NETCDF 3 (not compressed) size, in the same order than ZARR storage size

Generally greater in time extraction than Pangeo in CNES HPC, but faster for long term subsetting in CLS Cloud environment

Moving beyond the NetCDF TDS limits for global long-term analysis and modeling

To share a common storage among communities of users using different development/processing environments, cloud storage allows efficient R,Scala/Spark,Python/dask computing in memory with Notebooks.

To provide additional services:

to discover the data, describe the information for catalogues

to subset data if users want to download data on their premises

to enrich locations with environmental variables values and provide the relevant inputs for computing and modelling set up

Datasets, tests requests are available if to compare your solution/environment with our results (TileDB, COG…)

## Data Interoperability and Use Discussion

Robert Woodcock (CSIRO) summarized that the proposed interoperability BP development could include:

Cloud access/use for CEOS metadata services

Cloud formats BP

Cloud data lifecycle BP

Analytics interoperability

EAIL and JN BP

ARD strategy update

WGISS contributions.

# TECHNOLOGY EXPLORATION



## Introduction

Yousuke Ikehata (JAXA) introduced the session.

## [WGCapD Webinar Toolkit](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_WGCapD_WebinarToolkit.pptx)

Kenton Ross (NASA, WGCapD) discussed the CEOS webinar toolkit, the Jupyter Notebooks webinar as its prototype.

The toolkit includes how-to guidance, templates, and branding design assets. WGCapD also proposed a framework for CEOS working teams to gain WGCapD support during the webinar creation process and amplification of promotional efforts. Will include templates and design assets. Kenton described the toolkit structure, guides, templates, design assets, and reviewed the WGCapD support.

## [Jupyter Notebook for WGCapD Webinar](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_Webinar.pptx)

Esther Conway (UKSA) described the planned Jupyter Notebooks webinar, to be held July 21-22 with timings for two time zones. She gave the agenda, to include how Jupyter Notebooks can support CapD, JASMIN examples, SEO Google Earth, the Earth Analytics Interoperability, ESA - PGDS data cube example, and a panel. The target are training practitioners in the full range of agencies, and countries that want to deliver capacity development training.

There will be no space limitations. Publicity will be extended to WGCapD Networks and agency networks for CapD/ODA programmes.

## [CSIRO Jupyter Notebook Experience](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_CSIRONotebookExperience.pptx) [video link](https://eo-data.csiro.au/projects/WGISS-51/2021.04.21_TechExpo_CSIRONotebookExperience.mp4)

Matt Paget (CSIRO) gave a presentation on the CSIRO J-N experience. The key points are as follows:

Jupyter Notebooks and scalable data analytics

Details

* + Cloud-native platforms and EO archives
  + Advanced python tools designed for cloud solutions (Dask, Holoviz)

Challenges

* + Working with the cutting-edge software tools
  + Learning new patterns for lazy data analytics

Emerging outcomes

* + Data analytics patterns for ODC, Xarray and Dask and interactive visualisation with Holoviz
  + Positive experiences for scientists using cloud data & platforms

## [Developing a Jupyter Notebook Best Practice](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/2.Wednesday/2021.04.21_JupyterBP.pptx)

Esther Conway (UKSA) discussed developing a Best Practice for Jupyter Notebooks.

Key points:

Topic list and discussion points have been drawn up in the presentation

Some topics are more mature, others will require more discussion – How can this be facilitated?

A pragmatic approach is needed – consider publishing core topics first and then extending

Need to link the development to support of webinars and existing CEOS work

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

Esther listed the proposed content, and described the Binder repository deployment.

Rob noted that ODC Cube in a box and the SEO Google Colab ODC with Binder is an interesting option for some of the Jupyter Notebook BP binder deployment options

## Future Topics for Technology Exploration

Action WGISS-52-01: Yousuke Ikehata to initiate a discussion on what new topics WGISS should consider for exploration.

# Data PRESERVATION and STEWARDSHIP



## [CAL/VAL Maturity Matrix](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_CalVal_MaturityMatrix.pptx)

Paolo Castracane (Rhea, ESA) submitted a presentation about the Cal/Val Maturity Matrix. He listed the following progress:

* Easier and more focus on Cal/Val aspects.
* A coordinated, systematic framework for satellite mission quality assessment based on maturity matrix style reporting has been developed by ESA and NASA.
* Comprehensive assessment includes review of mission quality as evidence by mission documentation and detailed independent validation.
* Generic Assessment Framework implemented in detailed domain-specific guidelines. So far optical guideline drafted, work on SAR guideline recently initiated.
* NASA and ESA are trialling this framework in their current and planned commercial satellite data assessment/procurement activity.
* Such efforts have been on-going in recent years with commercial satellite vendors to understand and develop their quality information.

## [NOAA Maturity Matrix Self-Assessment Tool](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_NOAA_DataStewardshipMaturityQuestionnaire.pdf)

Nancy Ritchey (NOAA) discussed the NOAA Maturity Matrix Self-Assessment Tool. She reported the following:

Data Stewardship Maturity Matrix (DSMM) is a consistent framework for assessing and reporting quantifiable stewardship practices.

Manual DSMM assessment for many data sets is not sustainable.

Developed a questionnaire and implemented in the Collection Metadata Editing Tool (CoMET).

Scoring was determined by identifying every possible permutation of answer combinations per key component (>1600) and assigning a score to each.

DSMM results are included in the collection metadata record for each key component and the overall assessment and used in NOAA OneStop discovery interface.

Finalizing automatic generation of report to be preserved by the NOAA Central Library with cross-linking data and report DOIs.

Yves Coene asked if there is there a recommendation to include the DSMM as a standard way in the (ISO) collection metadata? Kent replied that yes, and there are some details provided on slide 15.

Esther Conway (UKSA) commented that it might be useful to provide guidance on how a mirror archive should cite a primary archive.

Action WGISS-52-02: DSIG to further elaborate on guidance on how a mirror archive should cite a primary archive.

Action WGISS-52-03: DSIG to organize a dedicated session on Quality Indicator and quality self-assessment (e.g., considering the work implemented in the Maturity Matrix)

Action WGISS-52-04: Michelle Piepgrass and Mirko Albani to see how WGISS can capitalize on NOAA’s CoMET.

## Persistent Identifiers

### [PID Open Points Discussion](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_%20Persistent%20Identifiers%20Discussion.pptx)

Mirko Albani (ESA) reported that there are nine discussion items / open points from previous meetings regarding Persistent Identifiers. These were discussed and disposed through the drafting of nine new recommendations to be added in the PID Best Practice.

The Data Integrity and Authenticity topic deserves more detailed discussion (e.g., on mechanisms for handling replica of data in the Cloud). Dedicated session will be held at WGISS-52.

### [NASA PID](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_NASA_PID.pptx)

Dawn Lowe (NASA) submitted a presentation on PID at NASA.

## Archive Holdings and Technology Session

### [ESA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_Archive%20Holdings%20ESA_DAS.pptx)

Daniele Iozzino (Rhea, ESA) gave a presentation on archive holdings and technology at ESA. She discussed current archiving technology and infrastructure, archive volume detailed per mission, operations concept, flows and processes. She also discussed data format/packaging for long term archive, management and archiving of the relevant associated information, and archive technology and media evolution and future trends.

Daniele identified the following challenges and needs:

Recovery of faulty tapes takes time. When a tape has I/O error, it has to be read from sequentially to re-archive all data contained

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 time

With the end of life of Storagetek T10K tape family and ORACLE HSM archive manager, it is feared that the main vendors are moving away from tape archive solutions. At the moment the only Enterprise solution is limited to IBM

### [NOAA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_NOAA_Data_Preservation-Stewardship.pdf)

Nancy Ritchey (NOAA) discussed archive holdings technology at NOAA-NCEI. She discussed the open data framework and gave details on the data holdings. They are dealing with challenges of moving to new technology while maintaining the legacy. She described the cloud archive workflow and technology and identified these challenges:

How to ensure sufficient citation and linking of assets?

Event 2 Thoughtful community best practices need to be developed and adopted

Ken commented that they are starting to focus on, the object level metadata. Every digital object should have a DOI, which stays with the object even if it moves to a different cloud provider.

### [STFC/CEDA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_DISIG_CEDA.pptx)

Esther Conway (UKSA) discussed archive holdings technology at UKSA/CEDA. She described details of their holdings and archive and archive services. She also discussed JASMIN, a facility operated by STFC on behalf of NERC. She described the archiving operations concept, flows and processes, and the data format/packaging for the long-term archive. She concluded with plans for future development.

Daniele Iozzino commented that their system is very similar with ESA’s, evolving their active to a mix of storage.

### [USGS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_USGS%20Archive%20Holidngs.pptx)

Ryan Longhenry (USGS) discussed archive holdings at USGUS. Their goals and objectives are to:

provide data management, access, archive, and distribution for all data sets within the USGS Historical Archives that have long term relevance to science and support the USGS mission,

improve access to the land archive,

utilize consistent data management approaches across all data sets,

develop and maintain data access, preservation, and distribution infrastructure to support the mission and other projects.

### [NASA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_ESDIS_Archive_Holdings.pptx)

Diane Davies (NASA) discussed the NASA/EOSDIS archive and data stewardship. She described the current archive composition and technology. She also discussed the preservation content implementation guidance. Diane concluded with a list of challenges:

* Ongoing efforts to address integrity of Archive: Considering checksums; e.g., use of sha256 instead of md5 and how this affects our collections and users. Ideas how to ‘watermark’ datasets to identify the origin of a dataset and ensure integrity
* Vendor Lock-in on the Earthdata Cloud to ensure that the collection can be moved from one Cloud vendor to another.

Usability and discoverability of data in Earthdata cloud. Capabilities for analysis ready data relative to a long-term archival format like HDF or NetCDF, and use of ZARR.

* Growth and changes to metadata: how to support old datasets with new metadata and enable them to be used by tools

Daniele commented that ESA is trying to address these challenges with blockchain technology. Andy added that NASA and ESA working together on this as part of their bilateral agreement.

### [JAXA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_JAXA%20Status%20on%20Archiving%20System.pptx)

Makoto Natsuisaka (JAXA) gave a status on archiving systems at JAXA. He described the archive policy for master data and products, and the current archiving system, technology and infrastructure. Makoto identified the following challenges and needs:

Archiving cost is serious problem and needs to be solved.

Distributed archiving systems among the JAXA centers might be needed to efficiently use computer resources. Introduction of JAXA DAB etc.

New technology like cloud, object storage might be needed to give flexibility.

A professional group for long-term preservation might be needed.

### [ISRO](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_ISRO_ARchiveHoldings.pptx)

Kalpana Tanguturu (ISRO) and Nitant Dube gave a presentation on archive holding and access mechanism at ISRO. ISRO has two major data holdings: NRSC and MOSDAC; the metadata is accessible to CEOS community. Nitant noted that ISRO has new data guidelines, with no limitations on commercial use. He described the access mechanisms, and gave details on the holdings.

## [International Cooperation on AVHRR Data](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_%20AVHRR%20International%20Cooperation%20Activities.pptx)

Mirko Albani (ESA) discussed international cooperation on AVHRR activities and partners.

The main issue is that archives not fully accessible event though open and free. The objective is to maximize the volume that are in these archives. Main actions:

Set-up a cooperative effort to unfold and make accessible 1km AVHRR data from regional archives (possibly open and free)

Transcribe unique data from heritage media

Identify a common format for AVHRR Level-1b data and pursue (re)processing from AVHRR data owners/holders into the commonly identified format

Facilitate data access through the WGISS Connected Data Assets Infrastructure

# Agency and Liaison Reports



## 

Agency reports were submitted by NASA, NOAA, CSA-CCMEO, CNES.

## [NASA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_NASAAgencyReport.pptx)

Andy Mitchell (NASA) reiterated that, as CEOS Chair there are a couple of chair themes are already being covered in WGISS. NASA will be reaching out to WGISS for support. NASA is promoting open source and now open science, and has agreements with Google, Microsoft, and Amazon, and is working with Google Earth engine and Google Cloud. 600 collections are on Amazon Cloud.

Please see details in the presentation.

## [NOAA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_NOAAAgencyReport.pptx)

Ken Casey (NOAA) reported that a lot of progress is being made with federal strategies Chief Data Officers, with an effort to make data a first-class citizen. They have wrapped up two big pilots in the cloud, and are doing some prototyping. They continue to make significant progress in moving data management and stewardship efforts to the cloud

NOAA is also leading a NOAA center for AI.

Please see details in the presentation.

## [CSA-CCMEO](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.01.22_%20CCMEO%20Agency%20Report.pdf)

Highlights:

OpenSearch: CWIC-ready RCM Open Search port to EODMS now completed (April 13, 2021)

Cloud: Presently engaged in several EO ecosystem integrations using microservices, OGC API, AWS cloud and Docker container technology – early focus on emergency response (flood mapping)

SBEO: Multi-department Space-based Earth Observation (SBEO) initiative in-progress to re-think EO value-chain from satellite to end-user

CEOS SAR ARD: Minor changes made in the metadata and requirements. Focus on avoiding unnecessary barriers for data providers or SAR missions. Changes are applied to both CARD4L-SAR Normalized Radar Backscatter and Polarimetric product formats. A revised version of the product format specifications will be submitted to the LSI-VC meeting (May 2021) for annual revision. CARD4L-SAR NRB/POL format will be tested in a cloud environment using reference data from Cambridge Bay. A tool is being developed work with CARD4L-SAR products

## [USGS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_USGS%20Agency%20Update.pptx)

The launch of Landsat-9 is in a few months, and the 50th anniversary of Landsat will be celebrated next year.

Regarding Sentinel/Landsat Data harmonization, USGS is working with NASA to make it visible; it will be found in the cloud in a couple of months.

Please see details in the presentation.

## [CNES](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_CNES%20agency%20report.pptx)

Please see details in the presentation.

# WGISS Plenary, part II



## [Future Meetings](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_FutureMeetings.pptx)

WGISS-52 will be held virtually held October 5-7, 2021 or October 19-21, 2021. This avoids conflicts with the 2nd Climate Observation Conference 12-14 October 2021 and IAC, 25-29 October 2021.

The consensus was for October 19-21, 2021.

WGISS-53 will be held in Spring 2022. Planning will be discussed in WGISS-52 since the COVID situation is unpredictable.

## [WGISS Summary and Discussion](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-51/3.Thursday/2021.04.22_Summary_Discussion.pptx)

Robert Woodcock (CSIRO) gave a summary of the meeting. He emphasized that a new vice chair for WGISS is needed, and requested nominations and volunteers.

Meeting highlights:

* Continue to see the evolution of STAC – an opportunity to fill gaps and revise information models.
  + Also, additional Quality, PID information and alternative ways of attaching them (Object level metadata in the Cloud)
* Seeing the advantages of Cloud processing and distribution
* Seeing the evolution of WGISS Discovery and Access to include the Cloud
  + Facilitating Discovery and Use (in-place not just access and download)
* Still work to be done to make Discovery, Access and Use easier
  + GEO common API for EO Data Cubes interoperability (any volunteers?)
  + ESA Vocabs and STAC vocabs standardisation? (And STAC standardisation!)
  + Replicas and Discovery in the Cloud

Damiano highlighted the last point and added that, regarding ARD, the focus is moving from format to architecture. A key point is whether to move the catalogs closer to the data, or having different catalogs that are federated. A parallel discussion is the handling of identity management.

[Kenneth Casey](res://\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20117%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%207667724%3c/onRightClick%3e%22%3e%3c/A%3e) commented that there is a continued and maybe growing appetite for common metadata authoring tools.

Potential direction for WGISS:

* PID and handling replicas in the Cloud (at WGISS-52)
* ARD Strategy update
  + Cloud Formats BP – Zarr needs work, netCDF in transition, COGS so far so good. Parquet is interesting
  + EO Vocabs for STAC – common names in STAC 🡪CEOS STAC BP
  + Further updates to ARD Terminology
* CEOS Cloud Data
  + At Scale things Fail
    - Cloud data management and notification BP to aid users
  + More on Cloud data discovery and evolution of WGISS IT
* Illustrate Use and Use in Cloud more:
  + Simple examples of using L1, L2 and L3 – deferring information loss to as late as possible
  + LIDAR & Data Cubes
  + Jupyter Notebooks BP – see the content outline in the presentation
* Supporting the CEOS Chair theme on Open Science and Decision Support
* Suggestions from SEO
  + Discourage downloading data and use data in the cloud​
  + Encourage the use of the cloud computing through CEOS prototypes (e.g., EAIL, ODC-Sandbox)​
  + Promote the production and use of ARD with STAC and COG formats​
  + Participate in Python notebook training events and focus on “simplicity”​
  + Demonstrate the value of radar data by working on ARD processing, data access, and training.
* Tech Expo – webinar topics and volunteers
  + Previous suggestions: cloud-native data formats, cloud data discovery and use (STAC), EO analytics on Kubernetes, cloud-based analytics architectures, Jupyter Notebooks for EO (in progress), blockchain

[Kristi Kline](res://\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20126%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%208257548%3c/onRightClick%3e%22%3e%3c/A%3e) noted that agencies are not all using the same cloud service provider; there needs to be emphasis on the ability to use data from different providers without having to move petabytes of data.

# Actions

Action WGISS-52-01: Yousuke Ikehata to initiate a discussion on what new topics WGISS should consider for exploration.

Action WGISS-52-02: DSIG to further elaborate on guidance on how a mirror archive should cite a primary archive.

Action WGISS-52-03: DSIG to organize a dedicated session on Quality Indicator and quality self-assessment (e.g., considering the work implemented in the Maturity Matrix)

Action WGISS-52-04: Michelle Piepgrass and Mirko Albani to see how WGISS can capitalize on NOAA’s CoMET.

# Glossary of Acronyms

AI Artificial Intelligence

API Application Programming Interface

ARD Analysis Ready Data

AWS Amazon Web Services

Cal/val Calibration and validation

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

DC data cube

DIF Directory Interchange Format

DOI Digital Object Identifier

ECV Essential Climate Variable

EO Earth Observation

ESIP Federation of Earth Science Information Partners

GCI GEOSS Common Infrastructure

GCMD Global Change Master Directory

GEO Group on Earth Observations

GEO-GLAM Global Agricultural Monitoring

GEOSS Global Earth Observation System of Systems

GFOI Global Forest Observations Initiative

GIS Geospatial Information System

GPM Global Precipitation Mission

GSDI Global Spatial Data Infrastructure

GUI Graphical User Interface

HPC High Performance Computing

IDN International Directory Network

ISO International Standards Organization

LSI Land Surface Imaging

LTO Linear Tape-Open

MOU Memorandum of Understanding

NRT Near real-time

OGC Open Geospatial Consortium

OSDD OpenSearch Descriptor Documents

PI Persistent Identifier

RS Remote Sensing

SEO Systems Engineering Office

SDCG Space Data Coordination Group

SIT Strategic Implementation Team

SLT System Level Team

SWG Standards Working Group.

TEP Thematic Exploitation Platform

ToR Terms of Reference

UML Unified Modelling Language

UMM Unified Metadata Model

VC Virtual Constellation

WCS Web Coverage Service

WG Working Group

WGCV Working Group on Calibration and Validation

WGCapD Working Group on Capacity Building & Data Democracy

WGClimate Working Group on Climate

WGDisasters Working Group on Disasters