MINUTES

OF THE

52nd MEETING

OF THE

CEOS WORKING GROUP ON
INFORMATION SYSTEMS AND SERVICES

(WGISS)

19 October to 21 October, 2021

Virtual Meeting

Table of Contents

[1 WGISS Plenary Session, Part I 5](#_Toc90192205)

[1.1 Welcome, Chair Report 5](#_Toc90192206)

[1.2 CEOS Executive Officer (CEO) Report 5](#_Toc90192207)

[1.3 Announcements of New Datasets, Services, and Tools 5](#_Toc90192208)

[1.4 Enabling Open Science 5](#_Toc90192209)

[1.4.1 Introduction 5](#_Toc90192210)

[1.4.2 OpenEO Platform: Current Status and Evolution 6](#_Toc90192211)

[1.4.3 Euro Data Cube - Bring Your Own Algorithm (BYOA) 6](#_Toc90192212)

[1.4.4 Open Science by Design 6](#_Toc90192213)

[1.4.5 USGS Supporting Open Geospatial Science 6](#_Toc90192214)

[1.4.6 Discussion 7](#_Toc90192215)

[2 Data Discovery and Access 8](#_Toc90192216)

[2.1 Status of GEO DAB Interoperability CEOS Catalogues 8](#_Toc90192217)

[2.2 General Reports 8](#_Toc90192218)

[2.2.1 IDN 8](#_Toc90192219)

[2.2.2 CWIC and CWIC Evolution 8](#_Toc90192220)

[2.2.3 FedEO 9](#_Toc90192221)

[2.3 Discovery of Services 9](#_Toc90192222)

[2.4 Processing-On-the-Fly Macro-Language (POF-ML) for Data Access and User Services 9](#_Toc90192223)

[3 Data INTEROPERABILITY and USE 10](#_Toc90192224)

[3.1 ARD Services for ECV Data in WGISS Carbon Community Portal 10](#_Toc90192225)

[3.2 Progress on developing ISO 19124-1 Calibration and Validation Standard 10](#_Toc90192226)

[3.3 Data Deployments in the Cloud – Summary 10](#_Toc90192227)

[3.4 Cloud Computing Technology 10](#_Toc90192228)

[3.4.1 Cloud Migration Archive Pilot 10](#_Toc90192229)

[3.4.2 Earthdata Cloud 11](#_Toc90192230)

[3.5 Update on Earth Analytics Interoperability Lab (EAIL) 11](#_Toc90192231)

[3.6 FDA and OSS Inventories 11](#_Toc90192232)

[4 Data Preservation and Stewardship 13](#_Toc90192233)

[4.1 Developing Community Guidelines to Promote Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets 13](#_Toc90192234)

[4.2 Term 'Level' Used in Maturity Matrix and Product Level Definition 13](#_Toc90192235)

[4.3 Session on Data Quality Self-Assessment and Indicators 13](#_Toc90192236)

[4.3.1 WGISS Data Management and Stewardship Maturity Matrix and Quality Indicators 13](#_Toc90192237)

[4.3.2 Data Quality at ESA 14](#_Toc90192238)

[4.3.3 Data Quality at NOAA 14](#_Toc90192239)

[4.3.4 Data Quality at JAXA 14](#_Toc90192240)

[4.3.5 Data Quality at USGS 15](#_Toc90192241)

[4.4 Provenance Using KSI Blockchain Pilot Project 15](#_Toc90192242)

[4.5 OAIS Interoperability Framework (OAIS-IF) 15](#_Toc90192243)

[4.6 Data Integrity and Authenticity on Cloud 15](#_Toc90192244)

[4.7 CEOS Common Online Dictionary 16](#_Toc90192245)

[4.8 International Cooperation on AVHRR Data 16](#_Toc90192246)

[5 Technology Exploration 17](#_Toc90192247)

[5.1 Webinar with Jupyter Notebooks Report with WGCapD 17](#_Toc90192248)

[5.2 Joint Webinar Planning with WGCapD 17](#_Toc90192249)

[5.3 Jupyter Notebooks Best Practice 17](#_Toc90192250)

[5.4 Blockchain KSI Technology 18](#_Toc90192251)

[5.5 NEODAAS AI Service or PML with AI and Machine Learning 18](#_Toc90192252)

[5.6 Technology Exploration Action Plan 18](#_Toc90192253)

[6 WGISS Plenary 19](#_Toc90192254)

[6.1 Agency Reports 19](#_Toc90192255)

[6.1.1 CONAE 19](#_Toc90192256)

[6.1.2 NASA 19](#_Toc90192257)

[6.1.3 USGS 19](#_Toc90192258)

[6.2 Future Meetings 19](#_Toc90192259)

[6.3 WGISS Summary and Discussion, Concluding Remarks 20](#_Toc90192260)

[7 Glossary of Acronyms 22](#_Toc90192261)

List of Participants

Agencia Espacial Mexicana Adrian Guzman, Julio Castillo

ASI Francesco Tataranni

CSA Mark de Jong (Canadian Forest Service)

CEOS Marie-Claire Greening

CNES Richard Moreno

CONAE Homero Lozza

CSIRO Robert Woodcock (WGISS Chair), Michelle Piepgrass (WGISS Secretary)

DLR Jonas Eberle, Karin Molch

ESA Mirko Albani, Priit Anton, Paolo Castracane (Rhea), Yves Coene (Spacebel), Philippe Goryl, Damiano Guerrucci, Patrick Griffiths, Daniele Iozzino, Grega Milcinski, Iolanda Maggio (Rhea), Filippo Marchesi, Stephan Meißl, Philippe Mougnaud, Javier Ojanguren, Sabrina Pinori (Serco), Serge Riazanoff, Roberto Roncella (CNR-IIA), Paulo Sacramento, Roberta Svanetti, Mattia Santoro, Anja Vrecko

EUMETSAT Joana Miguens, Michael Schick

European Commission Daniel Quintart

HSO Gábor Remetey-Fülöpp

INPE Lubia Vinhas

ISRO Nitant Dube, Sai Kalpana Tanguturu

JAXA Makoto Natsuisaka (WGISS Vice-chair), Yosuke Ikehata, Akihiko Kuze (WGCV Chair)

NASA Andrew Mitchell, Kathleen Baynes, Lauren Childs-Gleason, Valerie Dixon, Diane Davies, Dawn Lowe, Michael Morahan, Doug Newman, Ge Peng, Dan Pilone, Hampapuram Ramapriyan, Kenton Ross, Nancy Searby, Archibald Warnock, Min Wong

NOAA Kenneth Casey, Richard Baldwin, Aijun Chen (George Mason University), Prasanjit Dash, Liping Di (George Mason University), Gil Heo (George Mason University), Dana Ostrenga, Nancy Ritchey, Martin Yapur

NSMC/CMA Jinlong Fan

Polish Space Agency Oskar Zdunek

UKSA Esther Conway, Katie Awty-Carroll

USGS Tom Sohre, Cody Anderson, Peter Doucette, Steve Labahn

# WGISS Plenary Session, Part I

## Welcome, Chair Report

Robert Woodcock (CSIRO, WGISS Chair) welcomed the participants to the meeting, describing the logistics for this virtual meeting. Only a summary of each presentation will be given; the full presentations can be found on the meeting website (also linked to the title of each session).

## [CEOS Executive Officer (CEO) Report](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_10.05_CEOReport.pptx)

Marie-Claire Greening (CEOS Executive Officer) highlighted the current CEOS priorities, and the current Chair Theme, which is very relevant to WGISS:

Space-based Earth Observation Data for Open Science and Decision Support – and Priorities, including:

* 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) – CSIRO released the Earth Analytics Interoperability Lab (EAIL) which is now supporting several flood pilot projects.
* Open Data Cube Sandbox – The SEO released the Open Data Cube (ODC) Sandbox through 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. 12 application notebooks are available for users.

She reported that GEO is undergoing a mid-term evaluation, which is recommending that GEO evaluate and decide:

* what it wants or needs to pursue in terms of data infrastructure, producing data products, and user services,
* how GEOSS can integrate and execute the Knowledge Hub, and
* whether GEO has the capacity to carry this out.

Marie-Claire discussed the 2021-23 CEOS Work Plan Chair. For WGISS, under the category of Data Discovery, Access, Preservation, Usability and Exploitation (approaches, systems, tools and technologies), four deliverables remain open:

DATA-17-04: Technology exploration webinars and workshops

DATA-18-02: CEOS data holding reported and accessible in GEO and of her international relevant contexts.

DATA-19-02: Mekong Data Cube

FDA-19-01: Facilitate discovery and access for end users to data analytics and processing tools and services through the WGISS Connected Data Assets Infrastructure

Rob noted that the current remaining actions are all ongoing operations and require renewal/update as well.

## [Announcements of New Datasets, Services, and Tools](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_10.15_New%20USGS%20datasets%20and%20tools.pdf)

Tom Sohre (USGS) presented a USGS Dataset, Service, and Tool Update (since April 2021). News announcements of these can be found at: <https://www.usgs.gov/core-science-systems/nli/landsat/news>

Included are several Landsat collections, a new data access tool to Landsat, and collection 2 spectral indices and services. He concluded noting that Landsat 9 was launched successfully on September 27.

Rob commented that there has been heavy use of data and STAC API, which is a big advantage.

## Enabling Open Science

### [Introduction](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_10.20_Open%20Source%20Science.pptx)

Andy Mitchell (NASA) introduced the session for enabling open science via science data systems. The CEOS Chair theme’s intent is to not add to the significant number of commitments that CEOS already has. Instead, the concept is to broadly disseminate the relevance of CEOS remote sensing activities and contributions in support of Open Science and decision making.

Open Science and Decision Support is:

* Across communities, common themes characterize the concept of “Open Science” (e.g., availability, accessibility, transparency, collaboration).
* All of them align with increasing impact and accelerating science and decision support by lowering the barriers for all to contribute and benefit.

Open Science aims to ensure the free availability and usability of scholarly publications, the data that result from scholarly research, and the methodologies, including code or algorithms that were used to generate those data. (Reference: Open Science by Design, Realizing a Vision for 21st Century Research (2018))

Demonstrating Open Science with Purpose:

* Accessibility: call attention to CEOS datasets (free and open) and create more efficient access points (e.g., MIM, CWIC/FedEO, COVE) with connected training and capacity building.
* Transparency: promote sharing of data and algorithms among users and call attention to CEOS projects (e.g., EAIL, ODC Sandbox, COVERAGE) where this is demonstrated.
* Reproducibility: promote workflows (e.g., CARD4L, GEO Knowledge Hub) that support reproducible products where users can demonstrate open science
* Support Global Initiatives: demonstrate decision support for global endeavors (e.g., UN-SDG, GFOI, GEOGLAM) using an “open science” approach.

The goal of this session is to understand what open-science means to WGISS agency/program data systems.

### [OpenEO Platform: Current Status and Evolution](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_10.30_openEOplatform.pptx)

P. Griffiths (ESA) discussed the current status and evolution of the OpenEO Platform, which is a new ESA funded activity developing an operational service based on OpenEO API and a federation of cloud providers. OpenEO platform aims at addressing a prevailing capability gap and developing a service that corresponds directly to EO community requirements: enabling simplicity, Pixel to continental scalability, community and federation, providing transparency.

OpenEO platform offers users a Jupyter lab interface, the graphical web editor and links to QGIS. It supports client libraries in Python, R and JavaScript, and is being launched at the 2021 ESA Week and offers initially an Early Adopters and Training license package, with sponsoring from the ESA NOR.

Rob asked how the openEO Platform could support the CEOS priorities? or should it? Is the openEO API a data cube Interoperability solution? Is there a possible connection to WGISS Jupyter Notebook Best Practice?

Ken Casey (NOAA) wondered if an OpenEO platform is data equity towards ensuring data available to historically underserved populations. He added that his agency has begun talking about this, trying to be mindful. to break down historical biases and privileges, and wondered what others are doing in this area. Patrick replied that democratizing this work is central, providing effective ways to access and thus leading to equity; he noted that a stable internet is a requirement and as this becomes less of an issue it is worth pursuing.

### [Euro Data Cube - Bring Your Own Algorithm (BYOA)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_10.45_EDC_BYOA.pptx)

Anja Vrecko (Sinergise), P. Mougnaud (ESA) discussed the Euro Data Cube, which is fully operational. Its objectives are to get familiar with the EO data, explore, analyze and build and create applications. He stated that:

* 50+ PB of data available (Sentinel, Landsat, MODIS, Airbus, Maxar, Planet, Copernicus Services, ...)
* Apps, APIs and libraries for EO data processing
* Supporting 100s of applications and 100.000s of users worldwide
* 0.5+ BN requests processed monthly

Anja reported that Bring Your Own Algorithm was launched this week (PhiWeek 2021):

* It can deploy any Jupyter Notebook or Docker for headless execution
* It can execute on-demand via API or EDC Browser
* It is commercial and free options are available (support for billing)
* It is fully self-serving; support available if needed
* It is ready to be used
* Algorithms are needed!

### [Open Science by Design](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_11.00_Open%20Science%20By%20Design.pdf)

Katie Baynes (NASA) stated that Open Science builds on concepts from OpenSource Software revolution that expanded participation in developing code and applies it to the scientific process to accelerate discovery by openly conducting science from project initiation through implementation. She described the OpenSource Science Policy for the Earth System Observatory. To build such an ecosystem, one needs to:

* Initiate new missions, research and applied activities as open science projects.
* Implement clear policies for software, publications and data.
* Integrate and improve existing capabilities to support data management, access, computing, analytics and collaboration.
* Build the community through training, workshops, competitions and incentives

The focus areas are public engagement, capacity building, incentives.

Katie concluded saying that NASA has a strong vision for the future and has been making strides towards OpenSource science within the ESDS program and NASA Science more broadly. For upcoming missions, NASA will incorporate open science principles at project initiation to tackle common challenges. NASA is excited to continue to build a community dedicated to transparency, inclusivity, accessibility, and reproducibility.

### [USGS Supporting Open Geospatial Science](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_11.15_USGS%20Supporting%20Open%20Geospatial%20Science.pptx)

Pete Doucette (USGS) stated that the USGS approach to Open Science is an emphasis on maintaining science integrity and reproducibility. The impact to OGS has been best demonstrated via free and open data policy since 2008 and includes Landsat ARD, metadata, provenance and Level 2 (surface reflectance) processing algorithms. Potential investment areas are:

* Cloud access to Landsat data (Level 2 and Level 3); global ARD and Level 3 (e.g., LCMAP, NLCD, LANDFIRE, etc.) data and science algorithms.
* Open science algorithms for harmonization among multiple data sets (e.g., HLS).
* Promotion/collaboration with of Open Data Science and OpenAI platforms (e.g., Open Data Cube, Pangeo, other).

Potential investment areas to support open (geospatial) science into the future are:

* Cloud access to Landsat data (Level 2 and Level 3); global ARD and Level 3 (e.g., LCMAP, NLCD, LANDFIRE, etc.) data and science algorithms.
* Open science algorithms for harmonization among multiple data sets (e.g., HLS)
* Promotion/collaboration with of Open Data Science and OpenAI platforms (e.g., Open Data Cube, Pangeo, other….)

### Discussion

Andy Mitchell (NASA) led a discussion on Open Science.

Rob asked Katie how the Open Science movement fits with financially sustainable science? Some science organisations are under increasing pressure to create ROI. She replied that the agency is investing a lot, building tools that are reusable throughout the organization. for all science, not just earth science, but they are not focused on ROI at this point.

Andy asked Patrick if users can share their workflows with other users? Patrick replied that yes, visualized or exporter to json or translated to different programming languages; they are also considering sharing written code, or making repositories usable and accessible, and transition science algorithms to operational. Anja added that algorithm providers bring their algorithm to the platform; the provider can also open it and share it with others.

Rob asked if L3 data in the Cloud is needed, or simply published algorithms so users can process on the fly? Peter Doucette (USGS) replied that publishing L3, or earlier versions of data has been an interesting philosophical debate at USGS. I believe the future model will be to publish algorithms rather than data (especially multiple versions) to promote reconstitution of data in cloud environments as needed.

Tom commented that some organizations are replicating DCs, raising the question of how to know if they are authoritative or unchanged. Katie noted that they have researched into theoretical beyond checksum, working on data integrity as data is transformed but there is no good answer. Patrick added that there is no good solution from their side either. Tom said this is a good area for future research.

Nitant Dube (ISRO) asked if there is any framework to evaluate data quality from open science.

Priit Anton (Guardtime) noted there will be a solution and technology presentation tomorrow about data provenance. To answer the "show the origin of data though multiple user and supply chain".

# Data Discovery and Access

## [Status of GEO DAB Interoperability CEOS Catalogues](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_11.55_GEO%20DAB.pdf)

Roberto Roncella (CNR-IIA/ESA) discussed status of GEO DAB interoperability with CEOS catalogues, specifically with IDN, CWIC, and FedEO. He listed the following future enhancements:

* Avoid empty collections; sometimes opening a collection leads to zero results.
* Collection level OpenSearch responses provide a ‘Product metadata’ link which seems to contain additional information about the collection; this information could be used to enrich current mapping of information metadata. Sometimes these links return a JSON format instead of the declared XML format.
* Custom filters on granules search. Enhance second-level query capabilities by leveraging additional parameter.

Doug Newman (NASA) asked why split CWIC harvesting for CMR, since results will be exactly the same? Roberto replied that it is a policy decision. That will only work for CMR; there is no standard across your providers for granules-only.

Michael Morahan (NASA) asked what is meant by empty collections? No granules? Yes. Zero result for collection when we open it. It makes for a difficult user experience. This is valid for all CEOS systems. Michael said those collections can be filtered out, and that this should be discussed.

Damiano commented that the future enhancements are a good direction for better discovery of the data.

Yves Coene asked if they use only the collections that have an OSDD for product search?

Doug Newman commented that it would work but it would be better to filter them out in the query.

Yves Coene commented that the FedEO collections that have an OSDD attached for product/granules search should have granules. The other ones have no reachable granules currently.

Doug Newman noted that he would like to get a handle on how many of these collections exist. NASA has some but is working with providers to change that.

Yves Coene noted they are listed on the asset page which shows number of granules for each collection are available.

Roberto Roncella agreed to double-check and make a report on the "empty" collection for CEOS systems and try to improve this behavior.

## General Reports

### [IDN](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_12.00_IDN_Report.pptx)

Michael Morahan (NASA) gave a status report on the International Directory Network (IDN). He noted the ingestion of 21,609 NOAA dataset records, updates of 162 ESA, 53 DLR, and 44 VITO dataset records, and registration of five new ChinaGEOSS dataset records through dMMT. The following developments have occurred:

* Transition from Tagging to Metadata: OSDD implementation for IDN granule search now populated through Related\_URL fields
* Reminder to populate Data Usage/Licensing Policy using UseConstraints fields
* New associated\_DOIs field to cite related datasets.
* CEOS Tools records ingested into CMR for IDN <https://cmr.earthdata.nasa.gov/search/tools.umm_json?name=Proba-V%20MEP>
* Scheduling UMM-Services/Tools integration into Draft MMT
* Dynamic CMR STAC browsing support, e.g., [https://radiantearth.github.io/stac-browser/#/external/cmr.earthdata.nasa.gov/stac/LAADS/collections/MYD04\_L2.v6](https://radiantearth.github.io/stac-browser/)
* New GCMD Keyword Forum for requesting new/updates to Keywords [https://forum.earthdata.nasa.gov/app.php/tag/GCMD+Keywords](https://forum.earthdata.nasa.gov/app.php/tag/GCMD%2BKeywords)

Michael concluded with details on IDN usage metrics.

### [CWIC and CWIC Evolution](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_12.10_CWICReport_Evolution.pptx)

Minnie Wong (NASA) reported that the transition to CMR went well, and that ChinaGEOSS has been onboarded as a new CWIC data provider. Future plans are to examine the popular STAC standard as a means of federated discovery. In terms of the CWIC evolution:

* CWIC data partners have adopted and implemented recommended WGISS interoperable standard, [OpenSearch](https://opensearch.org/), and are encouraged to manage their own underlying connection to CWIC via tagging.
* CWIC initially developed as a separate software but functionality was transitioned to NASA’s Common Metadata Repository ([CMR](https://cmr.earthdata.nasa.gov/search)) in April 2021.
* The adoption of the CEOS OpenSearch standard and the use of International Directory Network ([IDN](https://idn.ceos.org/)) (which is built on top of CMR) by many CEOS agencies makes this possible.

### [FedEO](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_12.20_FedEO_Report.pptx)

Y. Coene (Spacebel/ESA) gave a status on FedEO. He reported that FedEO Client and Server have been rehosted under CEOS domain, documentation set updated (WGISS pages), and collections re-ingested in IDN. FedEO current metrics and DIF-10 export status include 94 to 107 million granules (+13%), with ongoing work with CCI, JAXA.

FedEO evolution activities include enhanced discovery interfaces, while preserving pre-existing interfaces. Yves reported that the ESE-ERGO project was completed May 2021with OpenSource components on GitHub (Metadata Editor and Metrics). He concluded with a EOVOC status update.

Roberto Roncella asked if the FEDEO service endpoint was updated? Currently in the GEO DAB they use the following: <https://fedeo.esa.int/opensearch/>. If the service endpoint is changed, GEO DAB need to update the endpoint and start interoperability tests.

## [Discovery of Services](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_12.30_Service_Discovery.pptx)

Yves Coene (Spacebel/ESA) gave a presentation on progress of discovery of services. The current “WGISS Connected Data Assets” made possible thanks to cooperation within WGISS/SLT the preparation of “CEOS OpenSearch Best Practices”, and the implementation of Best Practices by Partners and integration with IDN. ESA is seeking cooperation within WGISS/SLT on “CEOS Best Practices for Service Metadata and Discovery” to be elaborated within System Level Team (SLT). ESA is volunteering to prepare a working draft, aiming for inclusive work plan, including work already done.

Yves discussed the context, proposed work plan, and objectives and needs. He detailed specific use cases, discovery scenarios, and service metadata models. The current implementations are CMR, FedEO evolution, FedEO API evolution (RESTful).

Yves concluded with the following:

* Proposal to collaborate on “CEOS Best Practices for Service Metadata and Discovery”
* Draft Table of Contents
* Inputs (e.g., about more “current practices”) and contributions are welcome
* Objective to share (possibly through IDN) services/tools/applications described via agreed minimal metadata (model)

Doug Newman commented that it is very exciting to do this for services in the same way WGISS did for data. He added that, with the assumption that the WGISS solution to data discovery as successful, WGISS should capture the lessons learned from that activity and apply them to this endeavor formally.

Doug Newman highlighted that the main lesson from data discovery was 'use standards' but is sure there are more. The second takeaway was the provision of validation tools to assess a service's adherence to the various standards adopted. CMR is also leveraging something similar to rel=service in ATOM responses where collections are associated with tools and services. They have also been using Schema.org potential action elements to describe APIs to web tools such as Earthdata Search, State of the Ocean, etc.

## [Processing-On-the-Fly Macro-Language (POF-ML) for Data Access and User Services](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/1.Tuesday/2021.10.19_12.45_Processing-On-the-Fly_Macro-Language.pptx)

Serge Riazanoff (VisioTerra) gave a presentation on Processing-On-the-Fly Macro-Language. He discussed the following topics:

* Systematic / on-demand / on-the-fly processing:
* VtWeb – From high-level requirements up to design
* VtWeb – Design principles
* VtWeb – Data Processing Relay
* VtWeb – Generic platform for customized platforms
* VtWeb for geoservices

Serge also displayed exercises of

* Hello World of NDVI POF-ML
* Indicator for monitoring deforestation in Cameroon
* Cross Section Equalization (RCSE) applied to Sentinel-1
* Pan-sharpening of Sentinel-2 SWIR bands by its 10m bands
* Use of VtWeb POF-ML in CEOS / WGCV / TMSG / DEMIX

# Data INTEROPERABILITY and USE

## [ARD Services for ECV Data in WGISS Carbon Community Portal](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_10.05_ARDServiceForECVs.pdf)

Liping Di (NASA, NOAA) gave a presentation on ARD services for Essential Climate Variables (ECV) in the Carbon Community Portal. Liping mentioned that ECVs are key data sets for climate change studies. Analysis-Ready Data (ARD) are the data products 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 (defined by CEOS). Immediate analysis requires that data obtained by the data users exactly matches users’ specification in the format, projection, spatial/temporal coverage and resolution, and parameters so that it can be ingested into user’s analysis system immediately without further efforts. Since individual data users and projects have different requirements, personalized services for customizing the data must be provided in order to meet the requirement of immediate analysis, which we call ARD services. Most of ECV data providers do not provide such services.

CEOS WGISS developed a data portal for providing data discovery and access services to Carbon science community. All ECV datasets are accessible through the portal. ARD services have been prototyped for the ECV datasets discoverable and accessible through the portal

Liping detailed discovery and access of ECVs through the CEOS WGISS Carbon Community.

Rob asked if the use of WCS, people doing deep time series, how deep does it go in the WCS spec of Carbon Portal? Liping replied that the user can specify time series. If the user defines coverage with temporal domain, only one step at a time can be done, depending how the data is organized.

## [Progress on developing ISO 19124-1 Calibration and Validation Standard](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_10.20_ISO19124Report.pdf)

Liping Di (NASA, NOAA) discussed progress on ISO 19124-1 (calibration and validation standard). He noted that ISO TC 211 is an ISO technical committee responsible for setting international standards on geospatial information, including remote sensing information. ISO 19124 is a series of international technical specifications dedicated to calibration and validation of remote sensing data and derived products.

Currently ISO TC 211 is working on the ISO 19124-1 Fundamentals – Defining the common framework, methods, and UML structure – Allowing plug in other parts of the TS. CEOS, particularly WGCV, is heavily participating in and contributing to the project, which is currently in the working draft (WD) stage and expected to submit for Draft Technical Specification (DTS) in February 2022.

ISO 19124 has eight parts, including data from different types of sensors (optical, hyperspectral, lidar, SAR/InSAR, microwave), thematic products, and calibration/validation sites. Currently the project team (formed in December 2019i) is working on part 1; the first project team meeting was held in during ISO TC 211 49th Plenary in Japan in December 2019. Since then, multiple virtual project team meetings have been held to develop the TS – Large contributions from CEOS Cal/Val and WGISS, IEEE GRSS, and ASPRS; multiple people from CEOS are the project team members. Currently, WD version 0.9 has been released to the project team for review and comments.

Liping said he needs WGISS input to the report on ISO TC 211, and will contact WGISS in November.

## [Data Deployments in the Cloud – Summary](https://wiki.earthdata.nasa.gov/display/WGISS/CEOS%2BData%2BDeployments%2Bin%2Bthe%2BCloud%2B-%2BWGISS-52)

Diane Davies (NASA) reported that the team requested and received input from CEOS members on data deployments in the cloud. The information has been consolidated here in a table at this website: [https://wiki.earthdata.nasa.gov/display/WGISS/CEOS+Data+Deployments+in+the+Cloud+-+WGISS-52](https://wiki.earthdata.nasa.gov/display/WGISS/CEOS%2BData%2BDeployments%2Bin%2Bthe%2BCloud%2B-%2BWGISS-52)

The table displays a range of cloud providers; it is a living document, which hopefully will have information added over time.

## Cloud Computing Technology

### [Cloud Migration Archive Pilot](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_10.50_NOAA_NESDIS%20Cloud%20Archive%20Pilot%20%28NCAP%29.pdf)

Ken Casey (NOAA) discussed NOAA’s cloud migration archive pilot. He mentioned that NOAA/NESDIS began working in earnest to migrate to the cloud in 2019 following release of its 2018 [Cloud Computing Strategy](https://www.nesdis.noaa.gov/cloud-strategy). After successful completion of two pilots, NESDIS began operationalizing the NESDIS Common Cloud Framework (NCCF) in late 2020, and in 2021 conducted its NESDIS Cloud Archive Pilot (NCAP) to finish the end-to-end cloud archive prototype. It is preparing to operationalize this capability, with the goal of achieving an IOC by October of 2022.

Development of the NCCF in general and the cloud archive is being managed using the Scaled Agile Framework (SAFe), allowing for the coordination of multiple Agile teams. Included in the presentation is both architectural and schedule details for reference, to emphasize significant progress having been achieved but also to stress the amount of work remaining. The NCCF and the cloud archive component are being implemented within AWS, and the NASA CMR has been evaluated for possible incorporation into the NESDIS cloud solution (decision pending).

Ken concluded that the next program increments will focus on hardening of the demonstrated cloud archive features, and completing the remaining handful of features needed for the MVP. Resources are being aligned to support the cloud archive operationalization effort as early in 2022 as possible.

Andy asked about moving satellite control operations to the cloud. Ken said yes, but taking a wholistic approach. Activities are investigating how much more can be done in the cloud. NOAA as a whole being very aggressive in moving everything to the cloud, but mission critical satellite operations is farther in the future.

Homero asked how changing cloud provider can be done. Ken replied that they typically look across to other providers to ensure that those services are available elsewhere.

### [Earthdata Cloud](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_11.05_State%20of%20Earthdata%20Cloud.pdf)

Dan Pilone (NASA) gave a presentation on Earthdata Cloud (EDC), which enables Earth science data and computations to move into the cloud, creating opportunities for innovation around new services, such as sequencing data to support machine learning and application of large-scale analytics. Earthdata Cloud will improve the efficiency of data systems operations, increase user autonomy, maximize flexibility, and offer shared services and controls

Key End User Benefits are power, performance, freedom from data management, data co-location, and choice. Dan gave an overview of the Earthdata Cloud system. He also listed a number or lessons learned.

Dan mentioned the following near-term activities and goals:

* Continued migration of EOSDIS products and services
* Pilot-projects and movement towards Analysis Ready Data and more support for Analysis in Place
* Enhanced direct data access and simplification of user authentication / authorization
* Continued development and offering of “mix-and-match” services enabling easier data use for traditional users (e.g., sub-setters, format conversions, etc.)
* Enhanced metrics analysis capabilities
* Continued education and outreach efforts
* Additional “self-services” offerings for end users in controlled security boundaries
* Additional interoperability enhancements including continued expansion of STAC use

In response to a question from Andy, Dan said that standards are critical, and metadata is a key piece. One big problem is regional location of cloud providers. Co-locating data near each other helps but does not solve the problem.

Dan provided the following links:

* CMR STAC links: <https://cmr.earthdata.nasa.gov/search/site/docs/search/stac>;
* Earthdata Search link showing data available in AWS: <https://search.earthdata.nasa.gov/search?ff=Available%20from%20AWS%20Cloud>

## [Update on Earth Analytics Interoperability Lab (EAIL)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_11.20_EAIL.pptx)

Robert Woodcock (CSIRO) gave an update on the Earth Analytics Interoperability Lab. He began with background on the EAIL and reported that the EAIL as 46 registered users across CEOS projects and training is being provided to groups.

Strategy and next steps include:

* Sustaining the EAIL?
* Interoperability analysis: Specifically, the STAC API experienced, discovery via CEOS Services, and other CEOS Cloud data?
* EAIL and Jupyter Notebook BP

In terms of status:

The ROI and Products determined by CEOS projects include:

* USGS Collection 2 COGS Direct on AWS, STAC API
* Element 84 Sentinel 2 COGS: Direct on AWS, STAC API
* Sinergise S1 ARD Service: ordered-and-stored, fee for service
* EAIL S1 ARD Service: Native processing, EAIL stores ARD COGS
* NASA and COP DEM: EAIL stores copy

EASI Training notebooks EAIL tailored

Dask Gateway for scalable computing; Larger nodes, GPU – customisable

## [FDA and OSS Inventories](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_11.30_Inventory%20of%20OSS%20and%20FDA.pptx)

Iolanda Maggio (Rhea/ESA) reported that FDA and OSS Inventories accessible through CEOS web site need to be kept alive and up to date

Work in progress (in cooperation with SEO team):

* Administrator account to ingest new FDAs and/or OSS
* Ingestion tool improvement including cross check function (to identify broken links)

Next steps:

* Both inventories will be circulated (to update existing entries and/or add new ones if/when available) by end of October;
* Input to be received by end of November;
* FDA and OSS inventories updated online by end of December; inventories will be updated once per year.

Rob asked how WGISS can progress this area, given the work on services discovery, can it be used as a best practice? Damiano replied is this something that is feasible today to augment this storage of links with some additional metadata that would allow the discovery. Iolanda said there was an action at one time to put this in the IDN if possible. This inventory demonstrates how much OSS exists and wondered if there is a better way to make it (and its usage) visible. Damiano noted that the SLT could review this; Iolanda will work with them.

Michael added that the free and open ones were entered in the CMR but the only way to find the records is if the user knows them; they do have the keyword CEOS.

Yves noted that he is investigating if the spreadsheet can be aligned.

Rob added that the current work has served its purpose (in its current form), and suggested that the SLT work on a Best Practice of Services.

# Data Preservation and Stewardship

## [Developing Community Guidelines to Promote Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_10.00_FAIR-DQI_Guidelines.pptx)

Ge Peng (NASA) gave a presentation on developing community guidelines for sharing dataset quality information. She listed the following key points:

* Dataset quality is more than data quality. It includes quality of metadata, documentation, software, workflows, procedures, processes, and infrastructure, tools and systems.
* Dataset quality information is needed but hard to curate
* Sharing dataset quality information is desired but challenging
* Global Earth science community has asked for guidelines on sharing dataset quality information and a guidelines document has been baselined
* Improved sharing of dataset quality information takes a whole village

The International FAIR-DQI Community Guidelines Working Group comprises 22 global domain experts from government, academic, and private sectors. The development of guidelines is necessary since dataset quality information (DQI) is essential for users to trust, understand and effectively use the data, and curating dataset quality information is challenging (Jul 13, 2020 workshop (more details in the workshop report). Specifically:

* Multi-dimensionality of DQI; Fitness for purpose
* Lack of information; Across domain knowledge integration
* Limited and sparse community standards or guidelines
* Community feedback: Practical guidance is needed and should be developed by the community for the community!

Iolanda raised the point of the DMSMM, and the need to verify compliance. Peng replied that, when using DMSMM she would be happy to work with Iolanda on how to present the results. Mirko confirmed that they would like to interoperate on this activity.

## [Term 'Level' Used in Maturity Matrix and Product Level Definition](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_10.15_Terms%20in%20conflict.pptx)

Iolanda Maggio (Rhea/ESA) led a discussion on the term ‘level’ used in the Maturity Matrix and product level definitions. She noted that ISO19124: Geographic information — Calibration and validation of remote sensing data and derived products — Part 1: Fundamentals has potential ambiguity in the use of the term “level”. The WGISS Data Management and Stewardship Maturity Matrix is based on:

 - Components: Discoverability, Accessibility, Usability Preservation and Curation

 - Levels: from Not Managed to Well Managed

Conversely, the CEOS EO data Processing Levels (e.g., Level-1 data) also use the level terminology.

In each specific context there is no confusion, but using the term “level” in the same document with two different meanings might create confusion to the reader. She presented two solutions:

Option 1: Use another term in the maturity matrix like stage or degree

Option 2: apply a prefix in the maturity matrix (and ISO document) such as 'maturity-Level' or 'processing-Level’ (or apply only in the maturity matrix). This might lead to strange expressions like 'maturity-Level-X of processing-Level-Y data products’.

The consensus was to use option 2.

## Session on Data Quality Self-Assessment and Indicators

### [WGISS Data Management and Stewardship Maturity Matrix and Quality Indicators](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_10.20_%20DMSMM%20And%20Quality%20Indicators.pptx)

Iolanda Maggio (Rhea/ESA), P. Castracane (Rhea/ESA) gave a presentation on WGISS Data Management and Stewardship Maturity Matrix and Quality Indicators. She summarized with the following points:

* Overview on WGISS Data Management and Stewardship MM: Main Concept, Generation and Scope;
* Principles of the Quality Assessment Framework -> ESA/NASA Evaluation Framework;
* EDAP Best Practice Guidelines Activities -> Cal/Val Maturity Matrix;
* EDAP vs WMO vs WGISS -> Notes on possible evolution for WGISS SMM and formal link in the WGISS DMSMM White Paper;
* Way Forward: to implement the new proposed changes for the WGISS DMSMM improvement, insert in the WGISS DMSMM White Paper the annex with EDAP MM in order to give the opportunity to increase the quality check on the dataset measured. Finally, produce an automatic way to calculate the components measurement score.

Iolanda concluded with a Use Case of ENVISAT MIPAS L2 Data.

Paolo commented that in the last year ESA and NASA have launched an activity to evaluate data quality. This led to development of a common assessment framework, which is a work is in progress that depends on domain-specific matrices, with good progress for optical and radar.

To summarize, work by ESA continues to improve the WGISS MM, specifically with data quality. This work will be addressed at WGISS-53.

Rob commented that the use case in slide 17 would make a really good webinar in cooperation with WGCapD.

### [Data Quality at ESA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_10.30_ESA%20QualityIndicators.pptx)

Philippe Goryl (ESA) gave a presentation on data quality (quality indicators) work at ESA. He noted that the following aspects are recognized crucial for an optimal EO data quality assessment:

* The overall strategy shall be in line with the QA4EO principles
* Quality Indicators can be expressed in different layers of information:
	1. per pixel quality info (e.g.: cloudy, saturation, straylight, algorithm failed, etc.);
	2. Uncertainties per measurements (pixels/bands);
	3. Cal/Val results summary (geometry; radiometry; algorithms/methods and results compliance…) per data set;
	4. Summary quality in the metadata (importance of standardization);
	5. Tools and Procedures (including Product Generation and Quality tools)
* It is optimal to have quality information per pixel and at lower processing level
* Uncertainty characterization for EO Product shall follow standard guidelines (e.g. FIDUCEO and EDAP)
* Metadata shall include quality information and follow good practices and standards (e.g., CEOS Analysis Ready Data (ARD) metadata guidelines)
* Cal/Val activities shall be part of the overall Satellite Mission (see additional slides for Sentinel Data)
* Post-launch Cal/Val activities shall exploit appropriate infrastructure/methods traceable to SI (e.g., Fiducial Reference Measurements FRM).
* Validation plans are implemented involving: Mission Performance Centres and Validation Teams
* State-of-the-art algorithms, tools and procedures used, easily “fit for purpose” in terms of the mission’s stated performance.

He concluded saying that WGISS has a big role how to insert this information to the metadata.

### [Data Quality at NOAA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_10.40_NOAA-DataQuality.pdf)

Nancy Ritchey (NOAA) and Lihang Zhou gave a presentation on Data Quality at NOAA. She began saying that the challenge is to provide sufficient and consistent QA/QC information in a machine readable and actionable format, and curation of that information. NOAA has made recent progress on community driven efforts to develop QA/QC best practices, with consistent QA/QC and Cal/Val processes across NESDIS satellite programs, promoting use of best practices, and incorporating lessons learned and seek feedback from peers, users and stakeholders. NOAA also has extensive documentation of QA/QC and Cal/Val on Program websites, including some operational product monitoring (ICVS), and overall product QA/QC and maturity information is reported in the metadata. Data stewardship maturity assessments are utilized for some products.

Nancy reiterated what Ge Peng said: the data must be human readable AND machine readable. She concluded with details on tools and examples.

Mirko commented that comparing the two approaches is an excellent topic of cooperation with ESA.

Nancy added that the metadata includes: (each product) has a maturity level, and also the JPSS Algorithm maturity.

Rob noted that quality information is very context sensitive.

Philippe observed that that the cal/val MM is missing a fitness for purpose, and they are working to improve this aspect. The domain-specific scientist has to be involved at some level.

### [Data Quality at JAXA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_10.50_DataQualityatJAXA.pptx)

Makoto Natsuisaka (JAXA) provided answers to Iolanda’s questions: regarding data quality at JAXA:

Are you applying already defined Quality Indicators? Yes.

Do you have a study like QA4EO in place or finalized to address the topic?

Do you manage (and how) the Quality layer per pixels? Yes.

Do you manage (and how) the uncertainties per measurements? Yes.

Do you manage (and how) the Cal/Val results summary per data set? Yes.

Do you manage to put a summary quality in the metadata? Yes.

To perform all quality tasks are you using tools/apps? Data processing systems automatically generate quality indicators.

Makoto concluded saying that the details of the quality indicators are different depending on the satellites are displayed in the presentation.

Mirko remarked that it will be interesting to discuss summary quality information in the metadata at WGISS-53. Makoto said that he invited the chair of WGCV to attend. It will be good to have coordination between the two working groups at that meeting.

### [Data Quality at USGS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_11.00_USGS_DataQuality.pptx)

Cody Anderson (USGS) discussed data quality at USGS. He discussed:

* Landsat Collection 2 Per-Pixel Quality Bands: showed diagram discussing radiometric saturation quality assessment.
* Landsat Collection 2 CARD4L Assessment
* Landsat Per-Pixel Traceability, Measurement, and Measurement Uncertainty

He noted that mission-level information is also provided in Landsat Quarterly Quality Reports, Level 2 Validation, and System Characterization Reports

Mirko mentioned that a session for future meetings will cover the status of compliance with ARD specifications. Rob added that the ARD strategy update was done recently, and all working groups have a role to play in the implementation. A new coordination group being put in place and WGISS will need to have a representative. The strategy will be endorsed at Plenary.

## [Provenance Using KSI Blockchain Pilot Project](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_12.25_Provenance%20using%20KSI%20Blockchain%20Pilot%20project.pdf)

Priit Anton (GUARDTIME/ESA) gave a presentation on the Provenance Using KSI Blockchain Pilot Project.

* ESA and Guardtime are executing BC4SA-FO technology assessment project to demonstrate how KSI Blockchain will enhance security of data archives and deliver provenance.
* KSI Blockchain technology enables functionality, that provides integrity and provenance of satellite mission data (archiving and dissemination).
* Technology demonstration is provided on the ESA Space data preservation archive. New EO data security process is applied. Input EOSIPs are signed and verified to provide long term validation.
* Archive system operator receives overwatch capability of stored data products via dashboard. Evidence of integrity and provenance when distributing data products to third parties.

Priit concluded saying that this is very demanding (resource intensive), but it meets the standard for computation power. When asked about the algorithm used, he responded that hash is used though it depends on the desired security level.

## [OAIS Interoperability Framework (OAIS-IF)](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_11.30_OAIS%20Interoperability.pptx)

Roberta Svanetti (DEDACLOUD/ESA) gave a presentation on the OAIS Interoperability Framework. She reported that, starting from the initial concept contained in the OAIS Reference Model from the Consultative Committee for Space Data Systems (CCSDS) and the latest supplements in development by its MOIMS Data Archive Interoperability (DAI) Workgroup, interoperability between user, systems and archives has become necessary to guarantee and to improve exchange, understanding and use of the collections of information/holdings, preserved in the existing ESA archives from all disciplines, from Directorates, from external Parties.

Her presentation focused on the following key points:

* Standardization current status
* Concept and architecture status for the interoperability framework between user, systems and Archives
* Approach and scenarios adopted by ESA for the development of contextualized use cases in which interoperability and related standards lead to real benefits for ESA on its Heritage Data Repositories and Archives
* Planned next steps and potential cooperation in the frame of WGISS

She concluded saying that the EC is talking about standards and standardization and exchange data between domains and disciplines. The archives need to follow a common way to disseminate to make it useful. The future community of users is not yet known, so an infrastructure that is flexible and interoperable is used; standardization also improves the quality and the exchange between domains.

Mirko noted that ESA are moving ahead with interoperable pilot activities between archives, and encouraged others to contact them regarding this.

## [Data Integrity and Authenticity on Cloud](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_11.45_Data%20integrity%20on%20Cloud.pptx)

Iolanda Maggio (Rhea/ESA) gave a presentation on data integrity and authenticity in the cloud, in response to Action WGISS-49-13: WGISS DSIG to investigate how to handle replicas of data. She presented a table of mechanisms to handle a data replica on the cloud granting data integrity and authenticity.



Iolanda listed four approaches to manage Data replicas on cloud (to grant the Data Integrity and Authenticity) analysed:

1. Inclusion of a Persistent Identifier in the metadata file of each single product;
2. Inclusion of HASH Code embedded in any replicated and disseminated product;
3. Use of digital signature (e.g., Watermark);
4. Implementation of a Blockchain mechanism (e.g., KSI Blockchain)

Rob commented that this is valuable information that provides context for a conversation. Given the duplication that exists raises the following question:

* What does the user need to know?
* How does the agency protect itself?
* What is the danger with tampering – do the agencies have an understanding of this?

Andy noted that cloud is not much different than it was users got data directly from the agency, but that agencies can now take advantage of the fact that there is no longer a need for replication. NASA is trying to determine how to make the harvesters stand behind the agreements; one way is by having trusted sources. NASA and ESA are in a bilateral evaluation stage on this topic, along with NOAA.

## [CEOS Common Online Dictionary](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.00_CEOS%20Common%20Online%20Dictionary.pdf)

Katrin Molch (DLR) gave a presentation on CEOS common terminology. The vision is to establish a consolidated online source for EO/geospatial terminology to be used as the unique reference. A multi-disciplinary collaboration across Earth Observation and geoscience domains becomes more frequent and requires a common understanding of terms and definitions. A collaboration was initiated between CEOS WGCV, IVOS, and WGISS to “*advance the idea of a CEOS common online dictionary, with a view to eventually reaching out to launch a broader community effort (…)*” (WGCV Action Item 49-06, June 2021).

The status and way forward of this action is:

* CEOS Common Terminology group established
* Workplan drafted
* Merging WGISS Data Stewardship Glossary into CEOS WGCV ‘Terms and Definitions’ Wiki - ongoing
* Reach out to promote initiative - ongoing
* Discuss target infrastructure - upcoming

Katrin concluded with an invitation for anyone to participate. Mirko added that the initiative is very useful, noting that other communities have their own dictionaries that can also be integrated, such as LSI VC and Climate ECVs, and suggested to have a group to moderate in the event of disagreement.

Liping added that this would be a good collaboration with ISO TC211.

Ken commented that this would benefit from being in a wiki environment, and that NOAA is working on a lexicon, relying on existing terms whenever possible.

Mirko asked if this will be an action on the CEOS workplan. Katrin said it is inside the WGCV right now. Marie-Clare added that she would encourage adding something in the CEOS Work Plan. The WP is constantly under review, but the next WP (covering 2022-24) will eventually be finalized and endorsed by the next SIT in March 2022, so this is a good time to start to define and suggest new deliverable activities that will appear in the next WP update.

The scope at this time is CEOS only but including all EO and geospatial terminology can be in the future.

## [International Cooperation on AVHRR Data](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.15_AVHRR%20International%20Cooperation%20Activities.pptx)

Mirko Albani (ESA) discussed the international AVHRR data cooperation. He stated the following objectives:

* 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 data and pursue (re)processing from AVHRR data owners/holders
* Facilitate data discovery through the WGISS Connected Data Assets Infrastructure

The activities since WGISS-51 are:

* Inventory of existing national/regional HRPT and LAC data archives is under preparation. A template has been circulated, and feedback received from few people; more inputs needed. An email was sent to some regional archive owners to establish contact and involve them in the cooperation activities but no feedback has been received. Liaising will continue and assistance is welcome.
* Identification of high priority heritage media to be transcribed and approaches/opportunities: A pilot activity was y started by ESA with Memnon (Sony Europe) to assess feasibility of transcribing three types of Optical Disks containing unique AVHRR data: first results expected by end 2021. Data discoverability and reprocessing to be addressed at a later stage.

# Technology Exploration

## [Webinar with Jupyter Notebooks Report with WGCapD](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_11.55_JupyterWebinar.pptx)

Esther Conway (UKSA) gave a report on the recent webinar on Jupyter Notebooks, which was held in cooperation with WGCapD in July 2021. It was very successful, and there was significant participation. The recorded webinar can be found on the WGISS website.

Esther conducted a survey to gauge interest in future activities, and found that there is a lot of interest with hands-on training, and a lot of people interested in contributing to J-N training. The challenge is how to bring people together to do that. There is also interest in hack-a-thons.

Partially under development is a Best Practice; there is also a lot interested in contributing to BP and developing training materials.

Going forward, decisions should be made for the next training:

* What should be hands-on?
* How to include broad range of platforms and services?
* How to explain these services to potential users?
* How to target the right people for the right training?
* How to build on what already exists?
* How to best prepare potential users to exploit data?
* What to have resource for over the next year?

## [Joint Webinar Planning with WGCapD](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_12.00_WGCapD.pptx)

Kenton Ross (NASA) gave a presentation on WGCapD work planning. He listed the two relevant current Work Plan deliverables:

* [CB-20-06](http://deliverables.ceos.org/task_manager/deliverables/606/): Metadata Standards Tiger Team – Selection and definition of metadata standards for the exchange and processing of learning objects.
* [CB-20-12](http://deliverables.ceos.org/task_manager/deliverables/612/): Collaborative Feasibility Study (Mexico) – 10-week dual capacity building feasibility study focused on using Google Colab to create a flood monitoring algorithm in the Open Data Cube; working with CEOS SEO to explore the use of EAIL.

Other WGCapD interests include help with “infrastructure”: Metadata for learning objects, and, and a repository of learning materials.

Kenton listed the following WGISS-WGCapD potential collaboration points:

* Relevancy Ranking of Data Search Results
* Data Cubes for Large Scale Data Analytics
* The Burgeoning Role of Python for EO Data Analysis
* Explore Capacity Development with the Advanced Earth Analytics Interoperability Lab (EAIL)
* Joint Support for CEOS ARD – User Engagement, Pilot Activities, Expanded Uptake, Webinars
* Jupyter Notebook Best Practice Documentation/Training
* Follow-on Jupyter Notebook training activities with a Software Carpentry-style approach, could be a weeklong event with multiple threads
* Joint Webinars:
	1. The Open Geospatial Consortium (OGC) Coverage Standards Suite: Introduction & Overview
	2. Agile Development and the Scaled Agile Framework
	3. Future Data Access & Analysis Architecture Initiative Webinar (Joint WGISS/WGCapD Activity)

Kenton concluded with a description of their Webinar Toolkit Resources.

## [Jupyter Notebooks Best Practice](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_12.10_JupyterBP.pptx)

Esther Conway (UKSA) gave a presentation on the planned development of a Jupyter Notebooks Best Practice, which is a big opportunity for CEOS to build up capabilities. Esther listed the following topics/questions for the BP:

* Who and why to create a best practice document
* Notebook description and function
* Structure, workflow and documentation
* Technical dependencies and Virtual Environments
* Association with archived data
* Incorporation with data cubes
* Version control, preservation and archival, publishing software, getting a DOI and licence
* Interoperability and reuse on alternate platforms
* Creating a binder deployment

Esther suggested the following next steps:

* Detailed survey of training and Best Practice contributors
* Organize dedicated Jupyter Day for CapD training contributors April 2022 for full discussions
* Produce draft best practice of core topics for Jupyter Day
* Publish best practice on core topics after final review at WGISS - 54
* Organize Jupyter Week (hands on training) April 2023
* Develop advanced topics

## [Blockchain KSI Technology](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_12.25_Blockchain%20KSI%20technology.pdf)

Priit Anton (Guardtime/ESA) gave a presentation on KSI Blockchain technology, which enables functionality, that provides integrity, time and provenance of any digital assets. In space sector, focus is on satellite mission data (archiving and dissemination). It is globally accessible permissioned blockchain using widely witnessed events as proof to make the data already on blockchain impossible to change. It is in production since 2010. KSI Blockchain is built to digest data at massive scale and light interaction with legacy systems (push-pull requests and receive tokens, billions of requests in second)

A technology demonstration is provided on the ESA Space data preservation archive. Input EO-SIPs are signed and verified to provide long term validation. More info: <https://guardtime.com/eoguard>.

Priit discussed in detail the technology family tree, the KSI Blockchain Service, and technology differentiators. He also discussed differences between various Blockchain technologies (KSI, Bitcoin, Ripple, Hyperledger, NXT, Ethereum).

Yves Coene asked if it would be possible to agree on (custom) "metadata" in the NB Best Practice to be included in the Notebook file format itself (https://ipython.org/ipython-doc/3/notebook/nbformat.html) that can be extracted to catalogue/index the notebooks afterwards.

## [NEODAAS AI Service or PML with AI and Machine Learning](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_12.40_NEODAAS%20AI%20Service.pdf)

Katie Awty-Carroll (PML/UKSA) gave a presentation on the NEODAAS AI Service or PML with AI and Machine Learning (ML). She explained that the NERC Earth Observation Data Analysis and AI Service (NEODAAS) provides a range of services to EO researchers. There is increasing interest in AI for Earth Observation and research in this area is growing exponentially. However, research into AI for EO can be hindered by barriers to entry; to bridge this gap, NEODAAS has introduced a new AI service for NERC eligible researchers to get access to support for AI applications.

Through NEODAAS, researchers can access expertise in applying AI to EO data in addition to accessing our GPU cluster dedicated to EO. NEODAAS is already working on a variety of internal and external projects, including mangrove mapping, tree crown segmentation, and ship track detection. Future work will focus on expanding our training capability, developing robust pipelines, and researching solutions to common problems.

## [Technology Exploration Action Plan](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/2.Wednesday/2021.10.20_12.55_Technology%20Exploration%20Action%20Plan.pptx)

Yousuke Ikehata (JAXA) presented an action plan for future Technology Exploration discussions. He began by listing current activities, led by Esther Conway (UKSA): collaboration with WGCapD for Jupyter Notebook webinars (held July 2021), and developing a Best Practice for Jupyter Notebook.

The Technology Exploration Interest Group is tasked to investigate technologies in CEOS/WGISS that prove beneficial to the Earth observation community. The only topic being considered at this time is Blockchain. Yousuke requested volunteers to discuss this topic, and organize a session. He will begin to organize a monthly teleconference of the team, and invited members to join the email distribution list to begin this process.

Additional topics may also include ML (Machine Learning), DL, and AI (Artificial Intelligence). Rob agreed that AI is a topic that is still very relevant, and suggested reaching out to agencies each out to agencies.

# WGISS Plenary

## Agency Reports

### [CONAE](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.25_CONAE_Report.pdf)

Homero Lozza (CONAE) gave a presentation on recent events at CONAE. He discussed the SAOCOM project which enables Surface Soil Moisture (SM) monitoring at plot scale. The project promotes to add value by releasing open access products such as root soil water content, decision support systems for agriculture, and maps of anomalies and inter-seasonal mean yield scenarios for the main crops in the Pampas Region.

He concluded saying that a system for making mean yield scenarios maps based on normal or previous year estimates for wheat, maize and soybean is available. Three data access gates are available: GEOCatalog of metadata, GEOPortal and WMS GEOServices. Maps are updated weekly, and value is added to the SAOCOM mission with high level products which provide timely and valuable support information to stakeholders focused on the Pampas Region agriculture sector.

### [NASA](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.30_NASA%20Agency%20Report.pptx)

Andy Mitchell (NASA) gave a presentation on NASA agency updates. In summary:

* Delivered 1.9 billion products to over 4 million users in FY20 ((Oct. 1, 2019 to Sept. 30, 2020) to every country in the world. NASA Earth Observing System Data and Information System (EOSDIS) data have been distributed to 195 independent countries
* Provided data stewardship to almost 12,800 unique data sets for which we have minted over 9700 digital object identifiers that enable users to track data sets through publications and documentation.
* By the end of FY20, had archived over 41.5 Petabytes of Earth Science data from more than 150 instruments, available to users worldwide.
* Provided the ability to search over 34,500 data collections in the Common Metadata Repository, with 98% of queries completing in less than 1 second. The Common Metadata Repository itself has grown to manage over 710 million data records describing NASA Earth science data collection, a 40% increase over last year.
	1. Scored 79 on the American Customer Satisfaction Index (ACSI) survey of users (over 9000) with an increase of 3000 respondents over last year. This continues the trend of high scores for EOSDIS performance, despite the pandemic, a shift to working from home, and changes in our NASA missions and the research community schedules and plans.
	2. The LANCE system supports over 760 unique near real-time datasets. In FY20, LANCE produced 1.8 Petabytes of data within 3 hours of satellite acquisition that were distributed to over 650,000 users. 90.9 million files were distributed in FY20.
	3. In in 2020, Earthdata.nasa.gov website was visited by 436,402 people who visited/viewed over 1.7 million Earthdata information pages during the period. For the EOSDIS as a whole (including DAACs), we have 1.7 million people coming to our websites, viewing over 14.5 million pages of web-based information about Earth Science data and how to use it.

### [USGS](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.35_USGS%20agency%20report.pdf)

Tom Sohre (USGS) gave a presentation on recent events at USGS. He focused on the development status of Landsat operations:

* Landsat 7 flight operations is collecting about 470 new scenes per day; the science missing will end shortly after Landsat 9 becomes operational.
* Landsat 8 flight operations are collecting up to 740 new scenes per day, including frequent night and off-nadir imaging of volcano and fire.
* Landsat 9 is a near-clone of Landsat 8, but with important improvements for accuracy and resiliency. It was launched September 27.
* Landsat Next (late 2020s launch): NASA and USGS have set up formal projects to pursue Landsat Next. RFIs and instrument studies have been initiated.
* Landsat archive operations: over 9 million Landsat scenes are available, with 100 million downloads since Landsat became freely available in 2008.

## [Future Meetings](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.40_FutureMeetings.pptx)

Makoto Natsuisaka (JAXA) discussed plans for upcoming meetings:

* WGISS-53 will be held on March 21-24, 2022, in Tokyo, as a hybrid style meeting. This meeting will be held jointly with WGCV. Because of Covid-19, the final judgement is to be made in January 2022.
* CONAE kindly accepted to host of WGISS-54 in October in Buenos Aires.

Rob suggested contacting principals from each agency to see what their travel requirements prior to making the decision in January for WGISS-53.

## [WGISS Summary and Discussion, Concluding Remarks](https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-52/3.Thursday/2021.10.21_12.45_WGISS%2052%20Summary%20and%20Discussion.pptx)

Robert Woodcock (CSIRO) provided the following summary, with discussion topics, of the meeting. He remarked that this was a very production meeting, with excellent presentations, and thanked all the participants for their contributions.

CEO Executive Office report

CEOS Chair Theme links to WGISS:

* Data deployments in the Cloud
* EAIL and CEOS Data Cube
* Data Cube and interoperability (data and APIs)

WGISS Future actions?

* CEOS EAIL sustainability and expansion (More diverse nodes)
* Jupyter initiatives with WGCapD
* Support ARD strategy – specifics?

GEO Mid-term evaluation: reassess the concept of GEOSS – WGISS action?

Enabling Open Science

* Examples of accessing CEOS (Cloud) data in the more common APIs? 🡪 Jupyter Training target?
* WGISS BP in science reproducibility (at least the part it contributes)? 🡪 how to assure and cite CEOS data sources
* interoperable Data Cube APIs? 🡪 pick and promote a few common ones?
* Cal/Val Standard 🡪 Link explicitly to ARD strategy?

CEOS data and services in the cloud:

* Discovery expanding with Services BP (ESA lead), Cloud data (and replicas), and STAC browsing support: Continue yes, but generate some more specific actions for CEOS WP?
* Massive increase in CEOS data in Cloud, and used in EAIL, though more work to be done in interoperability, discovery, access, and use examples and also actual interop experience feedback. Cross-region – CEOS joint regional Caches?
* Work to be done on Data Assurance in Cloud

Data preservation and stewardship

* Sharing Data Set Quality
* WGISS Data Management & Stewardship MM: Well-aligned to Open Science Theme particularly reproducibility and quality.
* Data Integrity and Authenticity on Cloud: Next action worked out

CEOS Plenary Presentation

* Key outcomes:

Overall maturity and use of FDA, ARD markedly increased across agencies – now BAU?

Data in the Cloud: Availability and Discovery is now much broader: Cloud, Download, replicas, Services and Tools, Quality, Maturity

EO Jupyter BP and capability building with WGCapD

CEOS EAIL: More nodes and Interoperability experiments

* WGISS next steps

Embracing Open Science Theme

EAIL ongoing and expansion (Support EO Jupyter, evaluate interoperability and use with CEOS projects)

Continue to address upcoming challenges in Discovery (including services), Quality and Authenticity in Cloud

Further EO Jupyter BP

Continue growth of Data in the Cloud

# Glossary of Acronyms

API Application Programming Interface

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

DC Data Cube

DIF Directory Interchange Format

ECV Essential Climate Variable

EO Earth Observation

GEO Group on Earth Observations

GEOSS Global Earth Observation System of Systems

GIS Geospatial Information System

IDN International Directory Network

ISO International Standards Organisation

LSI Land Surface Imaging

OGC Open Geospatial Consortium

PI Persistent Identifier

PoC Point of Contact

SEO Systems Engineering Office

SIT Strategic Implementation Team

ToR Terms of Reference

VC Virtual Constellation

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