MINUTES

OF THE

49th MEETING

OF THE

CEOS WORKING GROUP ON   
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Virtual Meeting

Table of Contents

[1 WGISS Plenary Session, Part I 7](#_Toc40630306)

[1.1 Introduction 7](#_Toc40630307)

[1.2 WGISS Chair Report 7](#_Toc40630308)

[1.3 CEOS Executive Officer (CEO) Report 7](#_Toc40630309)

[1.4 Systems Engineering Office (SEO) Report 7](#_Toc40630310)

[1.5 CEOS Working Groups Reports and Progress on Cooperation with WGISS 8](#_Toc40630311)

[1.5.1 WGCapD 8](#_Toc40630312)

[1.5.2 WGDisasters 8](#_Toc40630313)

[1.5.3 WGCV-WGISS Joint Activities 8](#_Toc40630314)

[1.6 Discussion 8](#_Toc40630315)

[2 Data INTEROPERABILITY and USE 9](#_Toc40630316)

[2.1 FDA – Cloud based EO Analytics Architectures and Deployment/ Operations 9](#_Toc40630319)

[2.1.1 USGS Landsat Processing Cloud Architecture 9](#_Toc40630320)

[2.1.2 Open Data Cube on Kubernetes 9](#_Toc40630321)

[2.1.3 ESA Pixel Based Access and Data Cube Initiatives 9](#_Toc40630322)

[2.1.4 Status of the Brazil Data Cube Project 9](#_Toc40630323)

[2.1.5 Discussion 9](#_Toc40630324)

[2.2 User Experience and Portals for EO Analytics and Use 10](#_Toc40630325)

[2.2.1 Ghanaian Crop Modelling Jupyter Notebook and Data Cube 10](#_Toc40630326)

[2.2.2 JASMIN Jupyter Notebook Service Challenges and Opportunities 10](#_Toc40630327)

[2.2.3 Jupyter for CAPD – UK Overseas Development Aid, Agritech, etc. 10](#_Toc40630328)

[2.2.4 Discussion and Next Actions on Jupyter in EO 10](#_Toc40630329)

[2.3 Carbon Portal Update – Links to the Open Data Cube 11](#_Toc40630330)

[2.4 CEOS COAST - Users Needs and Experience 11](#_Toc40630331)

[2.5 Data Cube Interoperability 11](#_Toc40630332)

[2.5.1 openEO 11](#_Toc40630333)

[2.6 Data Interoperability and Use Discussion, Summary of Actions 11](#_Toc40630334)

[3 Data DISCOVERY and ACCESS 13](#_Toc40630335)

[3.1 General Reports 13](#_Toc40630337)

[3.1.1 IDN Report 13](#_Toc40630338)

[3.1.2 CWIC Report 13](#_Toc40630339)

[3.1.3 FedEO Report 13](#_Toc40630340)

[3.1.4 ESA Data Access Report 13](#_Toc40630341)

[3.2 SpatioTemporal Asset Catalog (STAC) 13](#_Toc40630342)

[3.2.1 STAC Overview 13](#_Toc40630343)

[3.2.2 Landsat and STAC 13](#_Toc40630344)

[3.2.3 ESA FedEO STAC 14](#_Toc40630345)

[3.3 Data Discovery and Access Discussion 14](#_Toc40630346)

[4 TECHNOLOGY Exploration 16](#_Toc40630347)

[4.1 Cloud Native Data Formats Overview 16](#_Toc40630349)

[4.2 Pangeo Ecosystem; xarray and zarr 16](#_Toc40630350)

[4.3 Update on CEOS ARD Production, Strategy and Approaches 16](#_Toc40630351)

[4.3.1 USGS 16](#_Toc40630352)

[4.3.2 NOAA: CoastWatch and OceanWatch ARD Preparation and Product Interoperability 16](#_Toc40630353)

[4.3.3 LSI-VC Update on CARD4L and ARD Strategy 16](#_Toc40630354)

[4.3.4 ARD and Data Cube for BIMSTEC 17](#_Toc40630355)

[4.4 Discussion 17](#_Toc40630356)

[5 Data PRESERVATION and STEWARDSHIP 18](#_Toc40630357)

[5.1 WGISS Data Management and Maturity Matrix 18](#_Toc40630359)

[5.2 NOAA CoMET Tool and the Maturity Matrix Self-Assessment Tool 18](#_Toc40630360)

[5.3 Persistent Identifier Implementation at CEOS agencies 18](#_Toc40630361)

[5.3.1 CEOS PID Best Practice 18](#_Toc40630362)

[5.3.2 ESA 18](#_Toc40630363)

[5.3.3 NASA 19](#_Toc40630364)

[5.3.4 NOAA 19](#_Toc40630365)

[5.3.5 JAXA 19](#_Toc40630366)

[5.3.6 CNES 20](#_Toc40630367)

[5.3.7 ISRO 20](#_Toc40630368)

[5.3.8 Technical Content and Information Preservation 20](#_Toc40630369)

[5.4 Discussion and Actions Review 20](#_Toc40630370)

[6 WGISS PLENARY, continued 22](#_Toc40630371)

[7.1 Future Meetings 22](#_Toc40630374)

[7.2 Discussion 22](#_Toc40630375)

[7.3 WGISS Summary 22](#_Toc40630376)

[7.4 WGISS-49 Actions 22](#_Toc40630377)

[7.5 Concluding Remarks 23](#_Toc40630378)

[7 Referenced Links 24](#_Toc40630379)

[8 Glossary of Acronyms 26](#_Toc40630380)

Australia Bureau of Meteorology Kate Roberts

CAS - AOE Xuesong Li, Ziyang Li, Chuanrong Li, Lingli Tang

CNES Richard Moreno

CONAE Homero Lozza, Josefina Otero

CSA Paul Briand

CSIRO Robert Woodcock (WGISS Chair), Cindy Ong (WGCV), Michelle Piepgrass (WGISS Secretariat)

DLR Jonas Eberle

EC Peter Strobl

ESA Mirko Albani (WGISS Chair), Yves Coene, Guido Colangeli, Andrea Della Vecchia, Patrick Griffiths, Guerrucci, Iolanda Maggio, Philippe Mougnaud, Marc Paganini

EUMETSAT Uwe Voges

GEOSS Steven Browdy

HSO, Ministry of Foreign Affairs & Trade Gábor Remetey-Fülöpp

IDEAM – Colombia Oscar Beltran

INPE Alber Sanchez, Gilberto Queiroz, Katyanne Conceicao, Lorena Santos, Lubia Vinhas, Michelle Picoli, Michel Chaves

ISRO Nitant Dube, Anupama Sharma, Indranil Misa, Pratyoosh SharmaRadha Krishna Kavuluru, Tushar Shukla, Saikalpana T

JAXA Makoto Natsuisaka (WGISS Vice-chair), Yousuke Ikehata

Natural Resources Canada Patrick King

NASA Andrew Mitchell, David Borges (WGDisasters), Lauren Childs-Gleason, Diane Davies, Valerie Dixon, Yonsook Enloe, Brian Killough (CEOS-SEO), Dawn Lowe, Christopher (Chris) Lynnes, Manil Makey, Andrea McCurdy, Michael Morahan, Douglas Newman, Mark Reese, Kenton Ross, Nancy Searby (WGCapD), Kurt Thome, Archie Warnock, Jennifer Wei, Min Wong, Tian Yao

NOAA Ken Casey, Rich Baldwin, Sid Boukabara, Paul DiGiacomo, Heng Gu, Veronica Lance, Dana Ostrenga, Ge Peng, Sathyadev Ramachandran, Nancy Ritchey, Kerry Sawyer (CEO), Michael Soracco, Martin Yapur

NOAA & NASA Liping Di, Eugene Yu, Aijun Chen

NRSCC Guoqing Li

UKSA Esther Conway, Rob Fletcher

UK Matthew Prior (CEDA), Richard Smith (CEDA), Ag Stephens (UKRI – STFC), Philip Kershaw (NCEO/CEDA) Jose Gomez-Dans (NCEO)

USGS Tom Sohre, Mathew Hanson, Kristi Kline, Steve Labahn

Planet Labs Chris Holmes

Sinergise Grega Milcinski

Columbia University Ryan Abernathey

Vienna Univ. of Technology) Matthias Schramm



Notes regarding this document:

Meeting website contains the agenda, with links to all presentations and documents:

<http://ceos.org/meetings/wgiss-49/>

Each section header of this document contains links to relevant presentations and reports.

Chapter 7 contains a table with follow-up URLs (web applications, demos, tutorials, and reports) mentioned at this meeting.

Chapter 8 contains a glossary of acronyms used

# WGISS Plenary Session, Part I

## Introduction

Robert Woodcock, WGISS-Chair (CSIRO) opened the WGISS-49 meeting, thanking everyone for their participation. He noted that this meeting is held virtually due to global circumstances, resulting in a packed agenda compressing the usual four full days into three four-hour sessions. Several technologies are available to enhance communication, and participants were encouraged to use them for their comments and other inputs.

## [WGISS Chair Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1140_WGISS%20Chair%20Report.pdf)

Robert Woodcock released his report (click on link in section title) summarizing WGISS achievements of the past year, discussing the approach to this first virtual meeting, and thanking members for all their contributions.

## [CEOS Executive Officer (CEO) Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1105_CEO%20Report.pptx)

Kerry Sawyer, CEOS Executive Officer (NASA) (click on link in section title) discussed the CEOS 2020-2022 Work Plan, outcomes of SIT-35 relevant to WGISS, including a specific request to WGISS (SIT 35-10) and key takeaways from the CEOS-GEO Bilateral Meeting relevant to WGISS.

SIT 35-10: The GEO Secretariat is requesting GEO Members and Participating Organizations nominate individuals to participate in the four GEO Working Groups (Capacity Development Working Group (CD-WG), Climate Change Working Group (CC-WG), Disaster Risk Reduction Working Group (DRR-WG), and the Data Working Group (Data-WG)) that were created under the *GEO Engagement Priorities Coordination* and *GEOSS Data, Information and Knowledge Resources* Foundational Tasks of the *2020-2022 GEO Work Programme*. Does WGISS wish to nominate a representative? Brian Killough will participate, but requests someone from WGISS to join; deadline is end of April.

Requests to CEOS from GEO:

* Produce data cubes as organized sets of cloud-optimized files with a catalogue; challenge because no agreed definition of DC
* Develop OSS tools for creating DCs
* Engage with OGC and GEO to conceive and develop web services for DCs
* Engage with OGC and GEO to design a common API for EO DCs and advance some measure of interoperability with CEOS and OGC (challenge though because of costs associated to change APIs)

**Note:** WGISS updates to CEOS Workplan due 10 May 2020.

**Action WGISS-49-01**: WGISS to identify members to participate in the four GEO working groups that were created under the GEO Engagement Priorities Coordination and GEOSS Data, Information and Knowledge Resources Foundational Tasks of the 2020-2022 GEO Work Programme.

## [Systems Engineering Office (SEO) Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1120_SEO_Report.pptx)

Brian Killough, Systems Engineering Office (NASA) (click on link in section title) discussed the status of the CEOS Open Data Cube (ODC), and developments of the CEOS Visualization Environment (COVE) tool. He presented the SEO’s new initiative to create an online cloud-based “Data Cube Sandbox” tool to support the Landsat Collection-2 release, and gave an update on data flows that make data available weekly; the SEO is working on automating the process.

Questions:

Chris Lynnes (NASA): Is there an API to COVE for machine-to-machine use? Yes, there is an API to COVE; Brian will review the best ways to advertise that.

Steve Labahn (USGS): Is the Element 84 S2 data in US-West AWS CARD4L-compliant? The ARD does not require a format, its requirements are more about documentation.

Nitant Dube (ISRO): Is there a tool for converting GeoTIFF to COG? Yes, there are many.

## CEOS Working Groups Reports and Progress on Cooperation with WGISS

### [WGCapD](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1130_WGCapD.pptx)

Nancy Searby (NASA), WGCapD Chair, gave a presentation (click on link in section title) on WGCapD activities. She listed past collaborative activities with WGISS, and presented the following opportunities for collaboration:

* Follow-on FDA webinar focused on specific FDA types
* CB-53: Collaborative Feasibility Study - 10-week dual capacity building feasibility study focused around data cube algorithm creation (focused in Vietnam, Mexico). Due: Q4 2020 or Q1 2021. WGCapD POC: Kent Ross, NASA, Kenton.W.Ross@nasa.gov

**Action WGISS-49-02:** WGISS, in cooperation with WGCapD, to develop FDA webinars for the following possible topics: cloud-native data formats, cloud data discovery and use (STAC), EO analytics on Kubernetes, ARD and ODC on Kubernetes, cloud-based analytics architectures, Jupyter Notebooks for EO.

**Action WGISS-49-03**: WGISS to determine how to achieve collaboration with WGCapD in the Interoperability Lab, such as a training facility.

### [WGDisasters](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1130_CEOS_WGDisaters.pptx)

David Borges (NASA), WGDisasters Secretariat, gave a presentation (click on link in section title) on WGDisasters activities. He highlighted the Geohazards Lab, an initiative to enable greater use of EO data and derived products to assess geohazards and their impacts –with a primary focus on hosted tools and services to exploit satellite data in cloud environments.

WGISS collaboration focus areas include:

* ARD Strategy / PFS
* Cloud Migration Good Practices
* Private Sector Engagement
* Determine joint Work Plan Deliverables for collaboration areas in 2021

**Action WGISS-49-04**: WGISS to determine joint WP Deliverables in collaboration with WGDisasters in the Interoperability Lab.

### [WGCV-WGISS Joint Activities](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1315_WGCV.pptx)

Cindy Ong (CSIRO), WGCV Chair, gave a presentation (click on link in section title) on WGDisasters activities. Cindy listed the four topics identified for cooperation between WGISS and WGCV at the joint meeting in April 2018 and highlighted the topic of CEOS Data Cubes and CEOS Test Sites Data Access in support of WGCV Activities.

Cindy also discussed the topic of Standardization and Best Practices (e.g. ISO 19159-3), whose objective is the contribution to the CEOS Best Practice/White Paper. WGCV’s collaboration with WGISS will complement the WGISS maturity matrix which focused on data management with quality aspects. The plan is to finalise discussion and presentation at virtual WGCV-47 meeting (July 2020), and to merge the CEOS WGCV/WGISS maturity matrixes at next joint WGCV/WGISS meeting.

Questions:

Liping Di (NASA/NOAA) asked about a new project – multiparty standards on calibration and validation for products. Cindy said would be interested in being involved.

**Action WGISS-49-05:** Liping Di to work with Cindy Ong on developing a project for multiparty standards on calibration and validation for products.

## Discussion

Nancy Searby suggested that several WGISS members join with WGCapD for collaboration. Robert pointed out that an upcoming session on the Jupyter collaboration should result in recommendations. He added that there is a proposal on the table about an analytics laboratory, around the issues, that is available [here](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/CEOS%20Earth%20Analytics%20Interoperability%20Lab%20Proposal%20WGISS%20and%20SEO.pdf).

# Data INTEROPERABILITY and USE



## FDA – Cloud based EO Analytics Architectures and Deployment/ Operations

### [USGS Landsat Processing Cloud Architecture](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1150_Data%20Interoperability-Landsat%20in%20Cloud.pdf)

Kristi Kline (USGS) gave a presentation on Landsat Processing in the Cloud (click on link in section title for the full presentation). Kristi gave a status update and described their infrastructure and interfaces, storage, data management, collection processing, and distribution and access. She noted that a STAC catalogue makes the data much more accessible and discoverable, and USGS is quickly moving to the cloud to process the collection. Kristi concluded with a discussion of egress use cases.

Questions:

[Robert Woodcock](res://\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20105%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%206881292%3c/onRightClick%3e%22%3e%3c/A%3e) requested discussion on possible ramifications for CEOS IDN, CWIC and FedEO given the changes to metadata and download options.

Homero Lozza asked about using OPeNDAP protocol for cropping areas or bands for download.

Robert Woodcock suggested that egress use cases should be discoverable information in CEOS WGISS discovery services so users know they can get better access if the move the computation to an appropriate Cloud provider-region.

**Action WGISS-49-06**: WGISS to determine joint WP Deliverables in collaboration with WGDisasters in the Interoperability Lab.

**Action WGISS-49-07:** WGISS to review impacts of cloud data availability and STAC in line with the changes at USGS.

Note: Nitant Dube suggested that WGISS work on a standard for retrieving data using a single interface irrespective of the formats in the manner done for OpenSearch and other standards.

### [Open Data Cube on Kubernetes](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1205_Open_Data_Cube_on_Kubernetes.pptx)

Robert Woodcock (CSIRO) gave a presentation on the Open Data Cube on Kubernetes (click on link in section title for the full presentation). He described Earth Analytics Science and Innovation (EASI) Hub with three layers of usage. He discussed the deployment into multiple availability zones, noting that the architecture can become challenging and complex. He also discussed cost optimizations (auto scaling), custom pods, security and reliability.

There was a request for a virtual workshop for ARD and ODC on Kubernetes.

### [ESA Pixel Based Access and Data Cube Initiatives](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1220_ESA%20Pixel%20Based%20and%20Datacube%20initiatives.pptx)

Philippe Mougnaud (ESA) introduced the Euro Data Cube (click on link in section title for the full presentation) and Grega Milcinski (*Sinergise*) gave a presentation it. The EDC has three components (OGC, Python, API) to make the data available. Grega described the data offer (from raw to ARD) and the EDC services (Sentinel Hub, Xcube, geoDB for vector data, Cloud resources).  He also discussed standardization activities and presented first use case for Land Parcel Identification with ML on top of EDC.

Andrea Della Vecchia (ESA) continued with a presentation on ESA pixel-based access and data access. The objectives are to bring the users to the data permitting pixel-based processing in a collaborative environment. They described the data offer and the authentication and authorization methods, and concluded with next steps.

### [Status of the Brazil Data Cube Project](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1235_BrazilDataCube.pptx)

Lubia Vinhas (INPE) gave a presentation on the Brazil data cube project (click on link in section title for the full presentation). She described the objectives and architecture, and continued with additional details, including the use of AWS.

[Steve Labahn (LSI-VC Co-Lead)](res://\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20127%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%208323084%3c/onRightClick%3e%22%3e%3c/A%3e) asked if they are considering adding Resourcesat-2 data to the data cube. Lubia replied that they are considering it, but it will depend on how long the agreement with ISRO will be maintained.

### Discussion

Robert Woodcock (CSIRO) led the discussion wondering what, with different types of metadata and storage locations, are the implications for the discovery methods. Andy commented that it can be an extension of the metadata, leading to possible enhancements.

Robert also raised the question of what CEOS can do for those agencies consuming data in DCs. Lubia replied that it took some time to develop the DC, but it is now quite stable and future versions are planned. STAC has made a huge difference. She added that the scripts are in GitHub, and they still have to improve the documentation to engage more people.

Ken Casey (NOAA) commented that adding something like STAC to existing services, not as a replacement but rather an addition is quite interesting. It would be very interesting to know what CWIC and IDN are thinking in terms of STAC. Integration of STAC catalogues deserves consideration at a formal level.

## User Experience and Portals for EO Analytics and Use

### [Ghanaian Crop Modelling Jupyter Notebook and Data Cube](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1100_Ghana_Data_Cube.pdf)

Jose Gomez-Dans (NCEO/UCL) gave a presentation on the Ghanaian crop modelling Jupyter Notebook and data cube (click on link in section title for the full presentation). He began with the background (monitoring agriculture in West Africa), describing the users, data requirements, and assumptions. He detailed the implementation of the data cube and centralized processing on JASMIN, and the user experience.

### [JASMIN Jupyter Notebook Service Challenges and Opportunities](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1115_JASMINJupyterNotebookServiceChallengesAndOpportunities.pptx)

Philip Kershaw (NCEO/CEDA) gave a presentation on JASMIN Jupyter Notebook Service (click on link in section title for the full presentation). He described JASMIN, a large-scale computing platform for data-intensive science, and its uses. He gave details on Jupyter and the evolution of JASMIN data analysis environments, and on Jupyter services in the context of JASMIN’s cloud and the evolution of data analysis environments. Philip described Cluster-as-a-Service behind the scenes and Notebook Service deployment. He concluded with early findings and thoughts for the future.

When asked who their users are, Philip replied that they are quite diverse among the environmental science community.

### [Jupyter for CAPD – UK Overseas Development Aid, Agritech, etc.](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1135_UK%20Overseas%20Development%20Aid.pdf)

Esther Conway (UKSA) gave a presentation on Jupyter for Capacity Development (click on link in section title for the full presentation). The objective is establishing effective coordination and partnerships among CEOS agencies offering EO education and training. She discussed how Jupyter might be useful to WGCapD, and why there is a need to collaborate with WGISS.

Esther discussed several examples. She suggested as a way forward a survey of agencies to learn targeted datasets and associated notebooks, and key datasets for WGCapD and associated notebooks. Also needed is input of relevant capabilities and needs and Notebook standards and CEOS endorsement.

### Discussion and Next Actions on Jupyter in EO

Robert Woodcock led a discussion on Jupyter in EO. Cooperation, interoperability and data integration need to be considered, as well as the role that WGISS will play. An interoperability laboratory is under consideration; are there other roles to consider?

Esther commented that there is potential data that could go up, especially data that does not lend itself to a data cube. Could it go there instead of being fragmented among agencies? There is enough data to test interoperability at different levels, without moving petabytes of data. There should not be a need to replicate data since agencies are putting it up.

Brian commented that three types should be considered: cloud, standard services, and new datasets that may need to be uploaded temporarily for study; this last is the worrisome one in terms of storage cost.

Philip Kershaw noted that clarification on the end goal would be helpful, as well as a roadmap of incremental steps.

Robert noted that a CEOS ARD and interoperability document is in draft. The primary driver is coming from activities like CEOS COAST (data along multiple boundaries), and describing levels that need to be addressed.

Brian reported that he looked through the IGARS suite of papers and could not find a single paper that used a combined set of satellite data; it is very difficult to combine. COAST will give the opportunity to bring multiple datasets together into a single system.

Ken Casey noted that there has to be further work with ARD to enable cross-domain interoperability.

Robert remarked that all these groups are aligning around DCs and ARD. Terminology needs to be precise. The interoperability lab will make it possible, and Jupyter will help bring that together

The central goal is to look at individual notebooks and work on issues of how we archive and use them.

With Jupyter Notebook there are potential risks around security. CSIRO is investigating two-point authentication and can report back.

## [Carbon Portal Update – Links to the Open Data Cube](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1200_CarbonPortal.pptx)

Liping Di (NASA/NOAA) gave an update on the Carbon Portal (click on link in section title for the full presentation). He described major activities and progress since WGISS-48, including links to the Open Data Cube. They have enabled metadata harvest from ODC at both collection and granule levels.

Eugene Yu (NASA/NOAA) gave a demonstration of the Carbon Portal.

## [CEOS COAST - Users Needs and Experience](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1215_COAST_WGISS.pptx)

Paul DiGiacomo (NOAA) gave a presentation on the CEOS Coastal Observations and Applications Study Team (COAST) (click on link in section title for the full presentation). COAST seeks to address the needs of the coastal community, cross cutting two domains. A key challenge is to bring together the systems and services that underpin the data. This is a transdisciplinary problem – ocean, terrestrial, atmospheric, and demographic. Ultimately the goal is to get more value out of Earth observations. Paul listed the project phases and deliverables, and concluded with next steps.

Liping noted a possible adaptation of the Carbon Portal to the COAST initiative.

## Data Cube Interoperability

### [openEO](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1230_openEO.pptx)

Matthias Schramm (Vienna Univ. of Technology) gave a presentation on openEO (click on link in section title for the full presentation). He described communication via openEO API, for easier communication between diverse communities and service providers. He discussed the interface and the virtual data cube view.

Nitant Dube asked if openEO uses any API gateway; is there capability for authentication and rate-limiting or does the client directly access the back-end. Matthias replied that the client need not access only the back-end. They are working on user authentication, but this will need funding, licensing.

## Data Interoperability and Use Discussion, Summary of Actions

Robert Woodcock (CSIRO) introduced the discussion commenting that with the changes to metadata and download options, there are ramifications for CEOS IDN, CWIC and FedEO. WGISS will support the current methods of searching data. However, the STAC records may be a better option with the transition to the cloud.

Jose Gomez-Dans commented that traceability is an issue to consider, as is harmonising the data sets.

Nitant Dube suggested that WGISS work on a standard for retrieving data using a single interface irrespective of the formats in the manner done for OpenSearch and other standards.

Nancy Searby noted that WGCapD can use help from WGISS in the area of data democracy, helping the less-involved agencies to bring their data into the mainstream, scaling up the engagement.

Thoughts regarding WGCapD/WGISS joint action on Jupyter notebooks:

Suggest refining the “awareness webinar” component of the action webinar and clarify that the objective of the webinar is to demonstrate how Jupyter notebooks to allow science and applications users to access and analyse Earth science data in architectures like date cubes and exploitation platforms.

Suggest a WGISS internal review to itemize Jupyter notebooks relevant to the Earth science community. The list of Jupyter Notebooks can be curated and shared. With these Jupyter Notebook instances in mind, interoperability can be considered.

**Action WGISS-49-08:** WGISS to coordinate an initial survey of agencies to learn targeted datasets and associated Jupyter Notebooks, and key datasets for WGCapD and associated notebooks. Survey could include input of relevant capabilities and needs and Notebook standards and CEOS endorsement.

**Action WGISS-49-09:** WGISS to break down the problem of defining the Jupyter Notebooks/Interoperability Lab to develop a clear idea of what this involves and where to go with it.

# Data DISCOVERY and ACCESS



Damiano Guerrucci (ESA) is the new lead of the Connected Data Assets System Level Team (CDA-SLT) upon the retirement on Yonsook Enloe (NASA) who served in WGISS since its inception.

## General Reports

### [IDN Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1310_IDN_Report.pptx)

Michael Morahan (NASA) referenced the IDN report (click on link in section title). He discussed CEOS/IDN collaborations and gave an update on the draft Metadata Management Tool (dMMT), commenting that the new dMMT Pathfinder functionality works well. He demonstrated recent IDN changes with the new Search Portal and gave highlights of updates with respect to GCMD, DocBuilder, and Search3 Portals.

### [CWIC Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1325_CWICReportFull.pptx)

Minnie Wong summarized an update of CWIC activities, noting that they plan to discontinue support of CSW. Please click on link in section title for the full report.

### [FedEO Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1325_FedEO_Report.pptx)

Damiano Guerrucci (ESA) gave a report on FedEO, noting the integration with the ESA thesauri, and listing the development plans through April 2021, including the adoption of OpenAPI, the adoption of new guidelines and standards, the transition to an internal metadata model and the development of a web-based metadata management tool. Please click on link in section title for the full report.

### [ESA Data Access Report](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1325_ESA_DataAccess_Report.pptx)

Damiano Guerrucci (ESA) announced the release on 27 April 2020 of the new Earth Online website which will be presented in a redesigned graphical interface with simplified navigation, mobile responsive design, powerful search functionality, and advanced filtering by thematic areas, missions, and instruments. ESA’s data exploitation is enhanced by providing a wide range of data offers through the main catalogue ([EO-CAT Web Client](https://eocat.esa.int/sec/)). Please click on link in section title for the full report.

## SpatioTemporal Asset Catalog (STAC)

### [STAC Overview](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1345_STAC%20Overview.pptx)

Chris Holmes (Planet Labs, Board of OGC) gave an overview of SpatioTemporal Asset Catalog (STAC) (click on link in section title for the full presentation). The main motivation for STAC is to avoid having to go to many different sites to find all the available imagery in an area, where every visualization shown is powered by a slightly different API.

Fundamentally, STAC is a language for geospatial catalogues and assets that focus on search and discovery. It is simple and extensible, with static catalogues and HTTP APIs. Chris gave an overview of the growing STAC ecosystem - a survey of tools and who is using them. The core is a set of tools to make and validate a desired set of catalogues.

One year ago STAC had one true content extension - the EO extension, as most participants were of satellite imagery background. Today there are extensions for SAR, point clouds and data cube content type, as well as ‘helper’ ones like checksum, date time range and scientific, for citations. The Label extension for machine learning came together at the last sprint.

STAC has always sought to align with WFS-3 (now called OGC API - Features).

### [Landsat and STAC](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1400_Landsat_and_STAC.pdf)

Matt Hanson (USGS) and Kristi Kline (USGS) gave a presentation of STAC use at USGS (click on link in section title for the full presentation). There is always a root catalogue that contains other catalogues and collections. Collections can contain other catalogues (called sub-catalogues) to group items further. Items are single scenes or set of data files for a specific location at a specific date and time. Links, link relation types, catalogues, collections, items (properties and assets), and content extensions were discussed.

### [ESA FedEO STAC](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/1.%20Tuesday%20April%2021/20200421T1415_ESA_FedEO_STAC.PPTX)

Yves Coene (ESA) gave a presentation of STAC use at ESA (click on link in section title for the full presentation). He discussed the work performed up to now (proof of concept, specifications, browse options and extensions), and presented plans for future work. He concluded that this is a partial implementation of STAC with drill-down via EO collections to EO products, co-existing with EO OpenSearch and OGC GeoJSON metadata specifications. He listed proposed extensions.

## Data Discovery and Access Discussion

Damiano noted that the presentations have a lot of details that are relevant, showing clear cases of STAC implementations. He asked Chris Holmes if agencies should wait for STAC v1.0 before beginning to use STAC. Chris suggested waiting for v1.0 beta1, but to begin getting ready with the existing version. Sometimes beta version is a better time to implement as it is stable and is so close to 1.0. It has the additional advantage that users can give feedback.

Damiano asked about the benefits of implementing STAC into the Landsat catalogue. Chris replied that USGS started using it just as they were migrating to the cloud, using Json files in the structure on the cloud. It has been a decent bit of work. When v 1.0 comes out they will put an API on it.

Makoto Natsuisaka requested more detail on the metrics, specifically user properties. Yves Coene replied that part of the user properties are obtained during the search, and parts later, so it is a combination.

Question: Is it premature to begin developing a best practice for the discovery processes that involve STAC?

Peter Strobl asked about common names in STAC how they are derived.

Chris Lynnes commented on the engagement with OGC: There are many (many) activities that overlap WGISS activities. The trick is to spend our precious time in the right ones.  A couple of the newer ones that may be of interest to WGISS are:

* [Environmental Data Retrieval API](https://github.com/opengeospatial/Environmental-Data-Retrieval-API) Standards Working Group - working on simplified APIs for extracting data in a variety of “shapes” (rectangles, curtains, trajectories, corridors) from an X-Y-Z-T field of data. (This is being driven by the meteorology community)
* [API - Coverages](https://github.com/opengeospatial/ogc_api_coverages):  an OpenAPI approach to WCS-like interactions

Chris Lynnes: Could the STAC proxy approach mentioned in Chris Holmes talk work with IDN et al? ESA FedEO STAC answers that question in the affirmative. IDN datasets whose metadata are hosted in CMR will be available via the STAC proxy approach (Chris Lynnes)

Peter Strobl: A question regarding the spectral band common names in STAC, how have these been created and should the concept not be made more generic and expandable when future missions arrive. (It seems very much focused on the status quo.) In essence this is an attempt to categorise the wavelength dimension in data acquisition, which is fair and welcome, but should follow the same good practices that apply for any discretisation of a domain.

Meaning it should be:

* unambiguous (no overlaps)
* comprehensive (no gaps)
* hierarchical (with some regularity regarding refinement levels)

For the optical wavelength range (usually 10nm to 100um) that would mean e.g. three main categories  at first level: UV, VIS, IR, then at second level: UV-C, UV-B, UV-A, Blue, Green, Red, NIR, SWIR, TIR, and so on, maybe each time dividing into three sub-categories and of course always with explicit wavelengths as limits. It needs also be based on clearly defined metrics “centre wavelength” needs to be defined properly with respect to the spectral response function (SRF):

* mean between min and max wavelength of SRF
* peak (maximum of SRF - lowest wavelength if several)
* centre of full width at half max (of real SRF or of Gaussian approximation)
* median of SRF

This should be done properly from the start as it is very confusing for users if these names change their meaning later on. Who would be the right people to discuss this with and how can we make sure to also get the standardisers (ISO, IEEE) involved?

**Action WGISS-49-10**: WGISS to discuss with LSI-VC the task of developing a list of common names in STAC how they are derived for consideration and feedback.

# TECHNOLOGY Exploration



## [Cloud Native Data Formats Overview](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1330_Cloud%20Optimized%20Data%20Formats_.pdf)

Chris Lynnes (NASA) gave a presentation on cloud-optimized data formats (click on link in section title for the full presentation). He discussed how data formats can enable high-speed access, and presented details of a cloud-optimized format study of classic and network-optimized formats. He concluded with summary of recommendations.

## [Pangeo Ecosystem; xarray and zarr](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1345_Pangeo.pdf)

Ryan Abernathey (Columbia University) gave a presentation on Pangeo (click on link in section title for the full presentation). He described Pangeo as a community platform for Big Data geoscience. The backbone of the Pangeo STAC is Jupyter Notebooks, xarray, DASK, and zarr. Pangeo allows representing big arrays as a virtual array of many small chunks: symbolic representation to do complex computations. The zarr library – is a new storage format: Open source library for storage of chunked, compressed ND-arrays. Ryan discussed Pangeo architecture and listed Pangeo deployments.

Ryan gave a demonstration (live demo reference in chapter 7) showing an interactive map streaming the data from cloud storage; a scale up computation by on-demand creating a compute cluster with rapid load of large amounts of data.

Ryan compared OPeNDAP with xarray + zarr, explaining how an API is not needed. He summarized as follows:

* Open source scientific python provides a great foundation for scalable earth system analytics (especially in the cloud).
* How to support / sustain open-source foundational software tools? (No agency or lab “owns” these, but they are critical infrastructure.)
* The best way to take advantage of the cloud is to give users direct access to analysis-ready data in object storage, not hiding it behind an API.

## Update on CEOS ARD Production, Strategy and Approaches

### [USGS](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1400_USGS_ARD.pdf)

Tom Sohre (USGS) gave a presentation on USGS ARD (click on link in section title for the full presentation). Tom discussed USGS and CEOS Analysis Ready Data for Land (CARD4L) efforts and described the USGS Landsat Collection 1 scene versus U.S. Analysis Ready Data tiled products. He described key USGS U.S. ARD product motivations and operational and science needs, Landsat Collection 2 data products.

Tom noted that USGS is in the final phase of reprocessing the consolidated Landsat global archive into a Landsat C2 data product inventory. Over the past two years, USGS has been scoping and working with the Landsat Science Team on defining and establishing a roadmap to potentially scale current U.S. ARD processing to the globe.

### [NOAA: CoastWatch and OceanWatch ARD Preparation and Product Interoperability](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1415_NOAA%20COAST%20WATCH.pptx)

Veronica Lance (NOAA) gave a presentation on NOAA CoastWatch and NOAA OceanWatch (click on link in section title for the full presentation). The goal of these projects is helping applications and missions that assess, understand, manage and protect ocean and coastal resources to benefit from satellite remote sensing observations. Veronica described NOAA CoastWatch, OceanWatch, and PolarWatch, and the value chain of the data, products, information and knowledge in ocean/coastal/inland water satellite observations. She displayed many examples of how these programs serve data products to its users, and discussed available “value-added” products/services. Detailed descriptions of the data portals were given.

Veronica concluded with future directions, challenges, and opportunities.

Chris Lynnes asked if this will be a key element of CEOS COAST. Veronica replied that they do hope to contribute.

### [LSI-VC Update on CARD4L and ARD Strategy](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1430_LSI-VC%20CARD4L%20Update%20and%20ARD%20Strategy.pptx)

Steven Labahn (USGS) gave an update on LSI-VC CEOS ARD for Land (CARD4L), and on ARD strategy (click on link in section title for the full presentation). He noted that USGS submitted Surface Reflectance (SR) and Surface Temperature (ST) self-assessments via LSI-VC in late 2019 for the CARD4Lself-assessments. The CEOS WGCV peer review team assembled and is evaluating submission. ESA and USGS are conducting self-assessments informing annual SR/ST Product Family Specification review/updates.

LSI-VC is actively involved in the CEOS Coastal Study Team (CEOS-COAST) with a strong interest in CARD4L products. This is an opportunity to integrate land, ocean, and atmosphere observations.

Steven also discussed the CEOS ARD strategy.

Robert Woodcock noted that AQUAWATCH is planning to collaborate.

### [ARD and Data Cube for BIMSTEC](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/2.%20Wednesday%20April%2022/20200422T1445_ARD_AND_DATACUBE_FOR_BIMSTEC.pptx)

D. Dhar (ISRO) gave a presentation on ARD and Data Cube for the BIMSTEC initiative (click on link in section title for the full presentation). The idea behind the initiative is to generate Landsat-8 OLI Datacube over the BIMSTEC region along with IRS data cube. The LSI-VC Team agreed to provide required data download and processing support and provided SIPG Team updated self-assessment document for CARD4L Compliance. He listed the activities and described their data cube. The Open Data Cube (ODC) team and ISRO team collaboration was discussed and put forward in motion at the latest LSI-VC telecon.

## Discussion

Philip Kershaw made the point that with the potential risk/power inherent with Jupyter Notebooks, security concerns need to be addressed if it’s to be used on a larger scale or in a shared way; e.g. built-in 2-factor authentication.

Step 0: WGISS Action 49-XX: WGISS to develop a white paper on native cloud data formats in the CEOS ARD strategy

Step 1: Action: Develop an ARD supply chain interoperability experiment

Step 2: WGISS Action 49-XX: WGISS to develop a white paper on CEOS ARD data supply chain Best Practices.

# Data PRESERVATION and STEWARDSHIP



## WGISS [Data Management and Maturity Matrix](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1100_CEOS%20WGISS%20DMSMM.pptx)

Iolanda Maggio (ESA) gave a presentation on the WGISS Data Management and Maturity Matrix (click on link in section title for the full presentation). She addressed various existing and new MMs and compared the WGISS, WMO, and EDAP MMs. Iolanda listed the major changes implemented recently, and discussed next steps and way forward.

Iolanda announced that the WGISS Data Management and Stewardship Maturity Matrix 1.2 document has been circulated to WGISS-Exec members. Integration of comments and input is expected by the end of May. By the end of June they hope to establish how to define levels assigning quantitative values instead of qualitatively.

Iolanda stated that thanks to this activity input can be provided to the RDA FAIR Maturity Model Working Group.

## [NOAA CoMET Tool and the Maturity Matrix Self-Assessment Tool](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1115_NOAA_CoMET_Tool_and_Maturity_Matrix_Assessment-Ritchey.pdf)

Nancy Ritchey (NOAA) gave a presentation on the NOAA Collection Metadata Editing Tool (CoMET) and the NOAA MM Self-Assessment Tool (click on link in section title for the full presentation). She described CoMET, showing record groups, dataset views, and rubric. Nancy also discussed the Data Stewardship MM guidance, template, tools, questionnaire, and questionnaire assessment. She concluded with plans for a DSMM report, and noted that NOAA is inserting overall ratings into the collection metadata record.

Iolanda asked if the questions are derived by the maturity matrix and if the tool provides an easy way to update questions? Nancy replied that the questions are derived from the MM, adding that the questions can be updated.

Mirko noted that the tool is to be used by data owners, but wondered if there are plans to release it to the public (specifically scientists). Nancy replied that yes, the scientists can do their own portion of the assessment but have to provide evidence that it is correct/valid.

Mirko asked of the tool is Open Source. Ken Casey replied that it is complicated; NOAA does not have an OS process or a clear formal policy, but the intention is to share it.

Iolanda suggested that DSIG be included in the list of assessors.

## Persistent Identifier Implementation at CEOS agencies

### [CEOS PID Best Practice](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1130_CEOS%20Persistent%20Identifier%20BP.pptx)

Iolanda Maggio (ESA) gave a presentation on CEOS Best Practices for Persistent Identifiers (PID) (click on link in section title for the full presentation). Iolanda described the focus of the BP document that is being written, and listed latest updates. Several open points were described.

Discussion:

Richard Moreno raised the point of whether there should be a new DOI for a copy of a collection in another format (e.g. zarr or COG). As a matter of fact, scientists will use this copy, and then will need to cite this recopied dataset, so there should be a new DOI. [Nancy Ritchey](res://\\G2MResource_en.dll/%3cA%20HREF=%22%3conLeftClick%3eeCMD_SetChatTo%20140%3c/onLeftClick%3e%3conRightClick%3eeCMD_DoAttendeeContextMenu%209175052%3c/onRightClick%3e%22%3e%3c/A%3e) added that the original DOI can be referenced in the new DOI to retain traceability. Richard and Nancy recommended that it should be a best practice.

[Iolanda Maggio](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) noted that the provenance of each dataset should be included in its landing page also.

Nitant Dube thanked WGISS team for writing the BP. ISRO has really benefitted from this when assigning DOIs to its products.

### [ESA](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1140_Persistent%20Identifiers%20for%20ESA.pptx)

Mirko Albani (ESA) gave a presentation on Persistent Identifiers at ESA (click on link in section title for the full presentation). Key points:

* The European Space Agency uses the Digital Object Identifier (DOI) system for its persistent identifiers.
* ESA DOIs are associated to collections and at a specific product processing level/product type or auxiliary files, and not to individual products/scenes.
* DOIs are assigned to data sets that are publicly disseminated and archived for the long-term – not to auxiliary data, experimental products, on-demand processing, or near-real-time products that are disseminated, but not archived.
* Currently, all ESA DOI related activities are performed manually; in the next few months an ESA DOI Service will be developed, in order to support all DOI related activities, to generate, register and manage DOIs. This will be a more automatized system.
* Each ESA assigned and registered DOI leads to a landing page, where information on the dataset collection and a link to access the data, is provided.
* DOI assignment and registration will continue throughout 2020 with targeted completion of assignment to all ESA EO data holdings.

### [NASA](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1130_NASA_PID.pptx)

Dawn Lowe (NASA) gave a presentation on Persistent Identifiers at NASA (click on link in section title for the full presentation). Key points:

* DOIs are assigned to NASA Earth science products that have long term usability, are likely to be cited, and are archived/distributed by EOS data providers.
* DOIs are assigned at the Collection level
* Each data product version is assigned a new DOI
* DOIs are assigned to documents relating to EOSDIS data products that may be cited in scientific publications – e.g., Algorithm Theoretical Basis Documents (ATBDs)
* DOI requests are processed by the EOSDIS DAAC responsible for archiving and distributing the digital objects
* Wherever applicable, the DOI is embedded within the digital object metadata
* ESDIS maintains a database of all project DOIs
* Process is largely automated
* DOIs are initially “reserved” to allow the DAAC to embed and test the DOIs in the product metadata before formal registration. Anything can be changed or deleted while reserved.
* The DOI resolves to a Landing Page, which in turn provides a link to the actual data represented by the DOI.
* DAACs working on establishing DOIs for older, less active data sets as they update metadata

### [NOAA](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1130_NOAA-Persistent%20Identifiers.pdf)

Nancy Ritchey (NOAA) gave a presentation on Persistent Identifiers at NOAA (click on link in section title for the full presentation). Key points:

* NOAA Data and Publication Citation Procedural Directive (PD) defines basis for Persistent Identifiers (Digital Object Identifiers) for a data collections
* Each DOI resolves to a landing page
* Planned work: Integrate DOI management into enterprise metadata tools
* Important to have a publicly-accessible policy statement, and to review and maintain it to keep up with and manage expectations, communicating updates and changes widely and quickly
* Do other repositories mint event-based DOIs for events?
* How should data citation information be embedded in distributed data files? Like a watermark for data.
* How can dynamic, custom citation information be provided to users for their ad hoc data access
* Support of DOIs required by publishers in order to publish articles. How are you handling that?

Andy commented that WGISS should consider the aspect of digital signatures; moving to the cloud DOIs will not solve it completely.

### [JAXA](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1130_PID_Implementation_JAXA.pptx)

Makoto Natsuisaka (JAXA) gave a presentation on Persistent Identifiers at JAXA (click on link in section title for the full presentation). Key points:

* ISAS/JAXA has already introduced DOI to a set of the space science data, and is planning to introduce them for Earth observation data.
* The draft plan is compatible with CEOS Persistent Identifier Best Practices except three points described below.
* The numbering will not be fully opaque.
* The information on the version will be described in the landing page and the citation document will include the version information.
* The information on the file format will be described in the landing page.
* Research products which are newly generated by investigators and have not been fully validated are basically out of focus for DOIs. At the same time those will attract much attention of many investigators. How are such products cited in publications?

### [CNES](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1140_Persistent%20Identifiers%20at%20CNES.pptx)

Richard Moreno (CNES) gave a presentation on Persistent Identifiers at CNES (click on link in section title for the full presentation). Key points:

* Two years ago CNES switched from PID at data collection level (ARK) to DOI at data collection level
* Follow CEOS PID BP with minor deviations
* Not all data collections have a DOI; waiting for new Data Management Strategy
* This year: change of CNES strategy for EO data management
  + All (past, future, present, external) projects will have a new DMP
  + DOI at data collection level
  + PID at granule level (including SHA256 hash code)

Mirko commented that the granule level assignment is interesting.

### [ISRO](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1130_Persistent_Identifiers_ISRO.pptx)

Nitant Dube (ISRO) gave a presentation on Persistent Identifiers at ISRO (click on link in section title for the full presentation). Key points:

* ISRO uses DOI for assigning Persistent Identifiers to its data products.
* DOI based Persistent Identifiers provide persistent citations in scholarly materials, this helps end users to retrieve metadata and data from the link available in the metadata. Long term actionable reference for satellite images enables discovery and re-use.
* All products available on MOSDAC have DOI assigned to them
* Each assigned DOI has associated landing pages
* ISRO plans to enable DOI based automatic discovery of references using web crawling

### [Technical Content and Information Preservation](http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-49/3.%20Thursday%20April%2023/20200423T1230_Technical%20Content%20and%20Information%20Preservation.pptx)

Iolanda Maggio (ESA) gave a presentation on preservation of technical content and information (click on link in section title for the full presentation). Iolanda announced that this will be a topic for the WGISS-50 meeting. She listed the topics that will be covered:

* Preservation processes
* Metadata and format definition and use
* Platform for long-term preservation (open source and/or licensed)
* Provenance and linked network
* Artificial Intelligence applications (e.g. natural language processing and machine learning to find reference to data in papers and link them to the ESA archives and to automate the keyword creation)
* Software and Tools Long term preservation

## Discussion and Actions Review

Mirko Albani (ESA) requested that questions and challenges on DOIs be sent to him in the next few weeks so that they can be considered and documented. He noted the following actions:

**Action WGISS-49-11:** Collect challenges/lessons learned/questions from WGISS members on PID/DOI usage/implementation (e.g. PID assignment to Document and SW by NOAA, Minor releases management, etc.) Collect from presentations and email from WGISS members.

**Action WGISS-49-12:** DSIG to organize a PID session to address/discuss questions and challenges and propose an updated version of PID Best Practice at WGISS-50.

**Action WGISS-49-13:** DSIG to investigate how to handle replicas of data for example in the cloud: PID assignment? Using watermarks? Mechanism for handling that to be further discussed and recommendation issued.

**Action WGISS-49-14:** DSIG to work on aBP on Data Associated Information management and preservation and discuss at WGISS-50.

Comments:

Nancy Ritchey raised two concerns regarding duplicate PIDs: exact replicas with multiple DOIs and transformations of the same data such as different formats.

NASA-ESDIS response to NOAA Questions:

* Event-based DOIs for cruises have been requested, but not adjudicated. Do other repositories mint DOIs for events? NASA-ESDIS is also discussing DOIs for events. We understand “cruise” event data to be a onetime data acquisition activity and can be treated similar to aircraft measurements. We are registering DOIs for aircraft measurements. We are discussing other types of “events" with respect to a special product created for an event e.g., Forest fire, earth quake, etc. which is like a dynamic product.
* How should data citation information be embedded in distributed data files? Like a watermark for data. NASA ESDIS has the requirement of including the data citation in the DOI Landing page. So by embedding the DOI reference in the data file's metadata, we achieve the goal of embedding the data citation information.
* Users access subsets of various data. How can dynamic, custom citation information be provided to users for their ad hoc data access? NASA ESDIS is discussing this under the topic of citing of data from tools and services.  The idea is having a DOI assigned to the tool/service with a way to capture the related DOIs for the input data collections and the call string containing the input variables used with that tool (subset parameters) and the resulting output slated to be referenced in a publication.
* Support of DOIs required by publishers in order to publish articles. How are you handling that? At NASA-ESDIS, each DAAC is responsible for upholding the NASA-ESDIS policy of DOIs being assigned only to data products slated for archival at a DAAC or managed by the DAAC. With that condition met, the DAACs can reserve DOIs or register a DOI for preliminary products for publication purposes.  If there is no prior agreement with the DAAC to maintain/distribute/archive the data product requiring a DOI, then that DOI request will be escalated to NASA ESDIS management for further consideration.

# WGISS PLENARY, continued



## Future Meetings

Makoto Natsuisaka (JAXA) reported that ROSCOSMOS has declined to host WGISS-50 due to COVID-19 concerns. WGISS-50 will be held virtually.

WGISS is considering a joint meeting with WGCV; this will possibly be WGISS-52.

## Discussion

Robert Woodcock (CSIRO) requested feedback on this first virtual WGISS meeting:

Positives:

* The meeting was conducted well
* Schedule was kept and timing of presentations was good
* High participation (as much as 60 people at once).

Negatives:

* Unable to discuss the issues and going to topics outside the agenda
* Having two in a row is not desirable and should consider postponing WGISS-50.
* Face to face is much better.

Suggestions for next virtual meeting:

* Use tools support breakout rooms for smaller groups with and video to allow for more conversations.
* Virtual coffee break.

## WGISS-49 Actions

Michelle Piepgrass (CSIRO) recorded the following actions:

**Action WGISS-49-01**: WGISS to identify members to participate in the four GEO working groups that were created under the GEO Engagement Priorities Coordination and GEOSS Data, Information and Knowledge Resources Foundational Tasks of the 2020-2022 GEO Work Programme.

**Action WGISS-49-02:** WGISS, in cooperation with WGCapD, to develop FDA webinars for the following possible topics: cloud-native data formats, cloud data discovery and use (STAC), EO analytics on Kubernetes, ARD and ODC on Kubernetes, cloud-based analytics architectures, Jupyter Notebooks for EO.

**Action WGISS-49-03**: WGISS to determine how to achieve collaboration with WGCapD in the Interoperability Lab, such as a training facility.

**Action WGISS-49-04**: WGISS to determine joint WP Deliverables in collaboration with WGDisasters in the Interoperability Lab.

**Action WGISS-49-05:** Liping Di to work with Cindy Ong on developing a project for multiparty standards on calibration and validation for products.

**Action WGISS-49-06**: WGISS to determine joint WP Deliverables in collaboration with WGDisasters in the Interoperability Lab.

**Action WGISS-49-07:** WGISS to review impacts of cloud data availability and STAC in line with the changes at USGS.

**Action WGISS-49-08:** WGISS to coordinate an initial survey of agencies to learn targeted datasets and associated Jupyter Notebooks, and key datasets for WGCapD and associated notebooks. Survey could include input of relevant capabilities and needs and Notebook standards and CEOS endorsement.

**Action WGISS-49-09:** WGISS to break down the problem of defining the Jupyter Notebooks/Interoperability Lab to develop a clear idea of what this involves and where to go with it.

**Action WGISS-49-10**: WGISS to discuss with LSI-VC the task of developing a list of common names in STAC how they are derived for consideration and feedback.

**Action WGISS-49-11:** DSIG to collect challenges/lessons learned/questions from WGISS members on PID/DOI usage/implementation (e.g. PID assignment to Document and SW by NOAA, Minor releases management, etc.) Collect from presentations and email from WGISS members.

**Action WGISS-49-12:** DSIG to organize a PID session to address/discuss questions and challenges and propose an updated version of PID Best Practice at WGISS-50.

**Action WGISS-49-13:** DSIG to investigate how to handle replicas of data for example in the cloud: PID assignment? Using watermarks? Mechanism for handling that to be further discussed and recommendation issued.

**Action WGISS-49-14:** DSIG to work on aBP on Data Associated Information management and preservation and discuss at WGISS-50.

## Concluding Remarks

Robert Woodcock (CSIRO) closed the meeting; he noted that this online/virtual meeting was very successful thanks to the available technology, and the excellent cooperation of all participants.

# Referenced Links

|  |  |  |  |
| --- | --- | --- | --- |
| **Org/topic** | **Speaker** | **Resource** | **URLs** |
| CEOS Earth Analytics Interoperability Lab | Robert Woodcock/ Brian Killough | Proposal | ceos.org/document\_management/Working\_Groups/WGISS/Meetings/WGISS-49/CEOS Earth Analytics Interoperability Lab Proposal WGISS and SEO.pdf |
| ODC GitHub | Brian Killough | Webapp | https://github.com/opendatacube |
| Data Cube User Interface Demo | Brian Killough | Demo | http://tinyurl.com/datacubei |
| Future Data Architecture Awareness Webinar | Nancy Searby | Webinar | <http://ceos.org/meetings/future-data-access-analysis-architecture-initiative/> |
| WGCapD Training Calendar | Nancy Searby | Calendar | <https://training.ceos.org/> |
| Best Practices for Conducting Trainings | Nancy Searby | Best Practice | <http://ceos.org/document_management/Working_Groups/WGCapD/WGCapD-Best-Practices_V2.0.pdf> |
| Ghanaian Data Cube | J. Gomez-Dans | Datacube | http://bitly/2VDUFzn |
| OpenEO | Matthias Schramm | Website | <https://openeo.org/> |
| JASMIN | Philip Kershaw | Website | [www.jasmin.ac.uk](http://www.jasmin.ac.uk/) |
| STAC | Chris Holmes | Specification | <http://stacspec.org> |
| STAC-ESA | Yves Coene | Rocket | <https://rocket.snapplanet.io/home?_url=https:%2F%2Fgeo.spacebel.be%2Fstac%2F> |
| STAC-ESA | Yves Coene | Browser | <https://geo.spacebel.be/?>[t=catalogs](https://geo.spacebel.be/?t=catalogs) |
| STAC-ESA | Yves Coene | API | <https://geo.spacebel.be/stac/> |
| Earth Online Website | Damiano Guerrucci | Website | [earth.esa.int](http://earth.esa.int/) |
| Cloud optimized format study | Chris Lynnes | Report | http://hdl.handle.net/2060/20200001178 |
| NOAA CoastWatch | Lance | Web app | [coastwatch.noaa.gov](http://coastwatch.noaa.gov) |
| Cloud Format Study | Lynnes | Report | <http://hdl.handle.net/2060/20200001178> |
| Pangeo | Abernathey | Website | http://pangeo.io |
| Pangeo | Abernathey | Web app | <http://gallery.pangeo.io/repos/pangeo-gallery/physical-oceanography/01_sea-surface-height.html> |
| Pangeo | Abernathey | Catalogue | <https://catalog.pangeo.io> |
| USGS ARD | Sohre | Data Format Control Book | <https://www.usgs.gov/media/files/landsat-analysis-ready-data-ard-data-format-control-book-dfcb> |
| USGS ARD | Sohre | Product Guides | <https://www.usgs.gov/land-resources/nli/landsat/landsat-science-products> |
| USGS ARD | Sohre | Website | <https://www.usgs.gov/land-resources/nli/landsat/us-landsat-analysis-ready-data> |
| CoastWatch | Veronica Lance | Website | [Coastwatch.info@noaa.gov](mailto:Coastwatch.info@noaa.gov) |
| CoastWatch | Veronica Lance | Website | <https://coastwatch.noaa.gov/cw/user-resources/satellite-data-training-courses.html> |
| LSI-VC CARD4L | Steven Labahn | CEOS ARD strategy | <https://ceos.org/ard> |

# Glossary of Acronyms

API Application Programming Interface

ARD Analysis Ready Data

AWS Amazon Web Services

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

COTS Commercial Off-the-Shelf

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

DC Data Cube

DIF Directory Interchange Format

DOI Digital Object Identifier

ECV Essential Climate Variable

EO Earth Observation

ESIP Federation of Earth Science Information Partners

GCI GEOSS Common Infrastructure

GCMD Global Change Master Directory

GEO Group on Earth Observations

GEO-GLAM Global Agricultural Monitoring

GEOSS Global Earth Observation System of Systems

GFOI Global Forest Observations Initiative

GHG Greenhouse Gas

GIS Geospatial Information System

GPM Global Precipitation Mission

GPU Graphics Processing Unit

GSDI Global Spatial Data Infrastructure

GUI Graphical User Interface

HPC High Performance Computing

ICT Information and Communication Technology

IDN International Directory Network

ISO International Standards Organization

LSI-VC Land Surface Imaging Virtual Constellation

LTO Linear Tape-Open

MM Maturity Matrix

MOU Memorandum of Understanding

NRT Near real-time

NWIP New Work Item Proposal

OGC Open Geospatial Consortium

PID Persistent Identifier

POC Point of Contact

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