MINUTES

OF THE

45h MEETING

OF THE

CEOS WORKING GROUP ON   
INFORMATION SYSTEMS AND SERVICES

(WGISS)

São José dos Campos, SP, Brazil

9 to 12 April, 2018

Hosted by

[Instituto Nacional de Pesquisas Espaciais](http://www.inpe.br/) (INPE)

Table of Contents

[1 WGISS Plenary Session, Part I 7](#_Toc512837971)

[1.1 Host Welcome, Review of Agenda 7](#_Toc512837972)

[1.2 WGISS Chair Report 7](#_Toc512837973)

[1.3 WGISS 2018-2020 Work Plan 11](#_Toc512837974)

[1.4 WGISS Brochure 15](#_Toc512837975)

[2 Data Preservation and Stewardship 17](#_Toc512837976)

[2.1 DSIG Standards Status and Way Forward 17](#_Toc512837977)

[2.2 Data Management and Stewardship Maturity Matrix Lessons Learned 17](#_Toc512837978)

[3 Data Discovery and Access 19](#_Toc512837979)

[3.1 WGISS Connected Data Assets 19](#_Toc512837980)

[3.2 International Directory Network (IDN) Report 20](#_Toc512837981)

[3.3 FedEO Status 21](#_Toc512837982)

[3.4 CWIC Status 22](#_Toc512837983)

[3.5 China (NRSCC) and China GEO Status 22](#_Toc512837984)

[3.6 Carbon Portal and Strategy and Demonstration 22](#_Toc512837985)

[3.7 OGC Documents Spec Status - OGC 17-003 and OGC 13-026r9 22](#_Toc512837986)

[3.8 CEOS OpenSearch Project Report 23](#_Toc512837987)

[3.9 Data Asset Enhancement — Synergies with FDA 24](#_Toc512837988)

[4 Data Interoperability and Use 25](#_Toc512837989)

[4.1 Recovery Observatory 25](#_Toc512837990)

[4.2 Open Source Software 25](#_Toc512837991)

[4.3 OGC Testbed 13 Outcomes and Information on the OGC Testbed 14 26](#_Toc512837992)

[4.4 User Metrics Introduction 26](#_Toc512837993)

[4.4.1 ESA User Metrics 27](#_Toc512837994)

[4.4.2 DLR User Metrics 27](#_Toc512837995)

[4.4.3 NASA User Metrics 28](#_Toc512837996)

[4.5 User management 28](#_Toc512837997)

[4.5.1 ESA Single Sign-On (SSO) and Federated Identity Management 28](#_Toc512837998)

[4.5.2 USGS User Management 29](#_Toc512837999)

[4.5.3 NextGEOSS User Management 29](#_Toc512838000)

[5 Joint Session with WGCV 30](#_Toc512838001)

[5.1 Welcome and Introductions, adoption of Joint Plenary Agenda 30](#_Toc512838002)

[5.2 Host Welcome Opening Address 30](#_Toc512838003)

[5.3 CEOS Executive Officer Report 30](#_Toc512838004)

[5.4 CEOS 2018-2020 Work Plan 31](#_Toc512838005)

[5.5 GEO Secretariat Report 31](#_Toc512838006)

[5.6 CEOS Systems Engineering Office Report 32](#_Toc512838007)

[5.7 WGISS Overview 33](#_Toc512838008)

[5.8 WGCV Overview 33](#_Toc512838009)

[5.9 Overview of WGISS/WGCV Joint Effort; Past Achievements & Discussion 34](#_Toc512838010)

[5.10 Data Formats and Interoperability in the framework of FDA 34](#_Toc512838011)

[5.11 Quality Indicators in Discovery Metadata 36](#_Toc512838012)

[5.12 CEOS Data Cubes and CEOS Test Sites Data Access in support of WGCV Activities 37](#_Toc512838013)

[5.13 Standardization and Best Practices (e.g. ISO 19159-3) 38](#_Toc512838014)

[5.14 Summary on Joint Interaction and Additional Opportunities; Joint Recommendations to CEOS and GEO 39](#_Toc512838015)

[6 Workshop on Future Data Architectures (FDA) and Introduction of Copernicus Data and Information Access System (DIAS) 42](#_Toc512838016)

[6.1 FDA Context 42](#_Toc512838017)

[6.2 Exploitation Platforms and Common Reference Architecture 42](#_Toc512838018)

[6.3 ESA Perspective on Data Cubes 42](#_Toc512838019)

[6.4 Sentinel-2 ARD: Making the Process Operational 43](#_Toc512838020)

[6.5 Open Data Cube in Uganda 43](#_Toc512838021)

[6.6 Horizontal Scalability for the Colombian Data Cube 44](#_Toc512838022)

[6.7 Copernicus DIAS 44](#_Toc512838023)

[6.8 GeoHazards Exploitation Platform (GEP) 45](#_Toc512838024)

[6.9 Proba-V Mission Exploitation Platform (MEP) 45](#_Toc512838025)

[6.10 USGS Experiences with Cloud Hosting 45](#_Toc512838026)

[6.11 CODE-DE Platform 46](#_Toc512838027)

[6.12 How Cloud Computing Can Help Scientists 46](#_Toc512838028)

[6.13 Cumulus: Cloud-based Archive for NASA’s Earth Science Data 46](#_Toc512838029)

[6.14 Best Practices for Cloud-Native Architectures 46](#_Toc512838030)

[6.15 NASA’s Earthdata Cloud Analytics Framework 47](#_Toc512838031)

[6.16 Vietnam Data Cube 47](#_Toc512838032)

[6.17 CNES FDA Status 48](#_Toc512838033)

[6.18 INPE’s MODIS Data Cube 48](#_Toc512838034)

[6.19 Summary of Workshop 49](#_Toc512838035)

[7 Agency/Liaison Reports 50](#_Toc512838036)

[7.1 CONAE 50](#_Toc512838037)

[7.2 CSIRO 50](#_Toc512838038)

[7.3 ESA 50](#_Toc512838039)

[7.4 Geoscience Australia 50](#_Toc512838040)

[7.5 INPE 51](#_Toc512838041)

[7.6 JAXA 51](#_Toc512838042)

[7.7 NASA 51](#_Toc512838043)

[7.8 UKSA 51](#_Toc512838044)

[8 WGISS Plenary Part II 53](#_Toc512838045)

[8.1 Future Meetings 53](#_Toc512838046)

[8.2 WGISS Summary 53](#_Toc512838047)

[8.3 WGISS-45 Actions 54](#_Toc512838048)

[8.4 Concluding Remarks 55](#_Toc512838049)

[9 Glossary of Acronyms 56](#_Toc512838050)

List of Participants \*via web conference

CAS/AOE Guangyu Liu, Hongyan He

CAS/RADI Yan Ma\*, Lizhe Wang\*, Jining Yan\*

CEOS Executive Officer Steven Hosford

CNES Richard Moreno

Colombia Data Cube Harold Castro\*, Pilar Lozano-Rivera\*

CONAE Homero Lozza

CSIRO Robert Woodcock, Cindy Ong, Matt Paget\*

DLR Albrecht von Bargen, Katrin Molch\*

EC Martin Ditter

ESA Mirko Albani (WGISS Chair), Yves Coene\*, Razvan Cosac\*, Andrea Della Vecchia, Damiano Guerrucci\*, Marco Leonardi\*, Rosemarie Leone\*, Cristiano Lopes\*, Iolanda Maggio, Philippe Mougnaud\*, Jolanda Patruno\*, Philippe Goryl, Michelle Piepgrass (WGISS Secretary)

GA Medhavy Thankappan

GEO Paola di Salvo\*, Nuno Catarino\*

GSDI/HUNAGI Gábor Remetey-Fülöpp

INPE Lubia Vinhas, Hilcea Ferreria, Karine Ferreria, Carlina Barrientos, Alessandra Gomes, Vitor Gomes, Thales Korting, Antonio Machado, Laercio Namikawa, Gilberto Queiroz, Joao Soares

ISRO Nitant Dube\*, Sai Kalpana\*

JAXA Makoto Natsuisaka, Akihiko Kuze, Yosuke Ikehata\*, Shinichi Sekioka\*

NASA Andrew Mitchell\*, Yonsook Enloe, Liping Di\*, Brian Killough\* (CEOS-SEO)\*, Li Lin\*, Chris Lynnes, Michael Morahan, Dan Pilone\*, Archibald Warnock\*, Eugene Yu\*

NOAA Martin Yapur\*, Anne Kennerley\*, Ken McDonald\*

NRSCC Xiaolong Dong, Chuang Liu\*, Junhua Ma\*

ROSKOSMOS Pavel Tischenko, Valery Zaichko\*, Tamara Ganina\*

UKSA Esther Conway, Nigel fox, Simon Reid\*, Federica Moscato\*

WGCV members Kurt Thome, Carolina Barientos, Fernando Camacho de Coca Xiaolong Dong, Nigel Fox, Philippe Goryl, He Hongyan, Akihiko Kuze Jamie Nickeson, Cindy Ong, Gregory Stensaas, Medhavy Thankappan, Albrecht von Bargen,



# WGISS Plenary Session, Part I

## Host Welcome, Review of Agenda

Mirko Albani (WGISS Chair, ESA) welcomed the participants to WGISS-45. He reviewed the agenda, highlighting the joint session with WGCV and the workshop on Future Data Architectures (FDA) and Introduction of Copernicus Data and Information Access System (DIAS). Mirko asked those present to introduce themselves.

Lubia Vinhas (Host, INPE) also welcomed the participants and reviewed meeting logistics, highlighting the group dinner planned for Tuesday night.

Michelle Piepgrass (WGISS Sec, ESA) reviewed the report from the WGISS Infrastructure Services Project (WISP). She described the project member roles and the WGISS organizational structure, and reviewed the method for submitting presentations. She also encouraged participants to supply the WISP team with all information related to email distribution lists and WGISS website contents. The principal purpose of WISP is to provide meeting support, and to assist with WGISS outreach activities.

## WGISS Chair Report

Mirko Albani gave a report on the CEOS and GEO highlights since the previous WGISS meeting, especially as they relate to WGISS.

Mirko began with highlights from the 31st CEOS Plenary meeting in October, where Robert Woodcock (WGISS Vice-chair, CSIRO) was confirmed as WGISS Vice-chair. He noted that the European Commission (EC) is the current CEOS Chair, which plans to ensure continuity and coherence of CEOS activities by:

* Supporting and further developing priorities and themes of the previous Chair (USGS)
* Understanding and defining common priorities with the incoming SIT Chair (NOAA)

Mirko added that the EC will work with USGS, ESA, NOAA and the relevant stakeholders to ensure continuity of the implementation of the study of future data access and analysis architectures, and supporting the ad-hoc team on Sustainable Development Goals (SDGs). The EC will work to ensure continuity of Moderate Resolution Sensor Interoperability (MRI). The CEOS Chair will also develop existing and future partnerships and priorities for CEOS, notably with GEO, the UN system, development banks, and the big data players, and expedite existing CEOS thematic acquisition strategies in relation to forests, agriculture, disasters, climate, carbon, and water. The CEOS Chair will focus on two initiatives, the second of which was detailed as it specifically affects the work of WGISS:

* Laying the foundation for an international CO2 and GHG emission monitoring system
* Bringing the benefits of Future Data Architectures to the present; identify new targets

Mirko reminded of the upcoming SIT-33 meeting where WGISS will participate, and which will focus on two of the 2018-2019 SIT Chair priorities:

* Address Working Group (WG) and Virtual Constellation (VC) continuity, sustainability and outputs, including seeking observations from VCs and WGs on best practices and possible modifications to existing practices. Identify synergies and cooperation areas between VCs and WGs.
* Improve and clarify CEOS relationships with CGMS, GEO, and WMO by identifying coordinated activities and interaction where appropriate, and at the same time map CEOS priorities with the 2017-2019 GEO Work Programme.
* Coordinate an FDA workshop, with objectives to examine past and current CEOS agency initiatives involving commercial ICT providers, their integration into the service offerings of CEOS agencies, and the way in which these facilitate the development of value-added services.

Mirko described aspects of the CEOS 2018-20 Work Plan which includes 12 actions assigned to WGISS, and two actions where WGISS will provide support. WGISS will review the status and progress of these actions, which are:

CARB-15: Carbon Data Portal prototype – Q3 2018 Ken Casey and Ken McDonald (NOAA)

FDA-5: Promote awareness of FDAs (support WGCapD) – Q4 2018. Mirko Albani (ESA), Robert Woodcock (CSIRO), Steven Hosford (CEO), Richard Moreno (CNES).

FDA-8: Establish a common description of Future Data Architecture functional blocks and identify interfaces and interoperability approaches. (Support FDA AHT) – Q3 2018. Mirko Albani (ESA), Robert Woodcock (CSIRO), Steven Hosford (CEO), Richard Moreno (CNES), Chris Lynnes (NASA).

FDA-9: Inventory and characterise existing FDAs operated by both public and private entities including the standards and approaches they use (e.g. Data Cubes, Exploitation Platforms, Copernicus DIAS, etc.) – Q3 2018. Mirko Albani (ESA), Robert Woodcock (CSIRO), Steven Hosford (CEO), Richard Moreno (CNES), Chris Lynnes (NASA).

FDA-10: Finalise inventory of Software and Tools available or used at CEOS agencies for EO data exploitation and use focusing on Open Source but remaining as broad and inclusive as possible and implement a mechanism for discovery and access – Q3 2018. Iolanda Maggio (ESA) and Robert Woodcock (CSIRO).

FDA-11: Organise several sessions/workshops to share lessons learned and outcomes from FDA systems and platform pilots and Interoperability Projects. With FDA AHT – Q4 2018. Mirko Albani (ESA), Robert Woodcock (CSIRO), Steven Hosford (CEO).

FDA-13: Develop a User Metrics Best Practice – Q4 2018. Mirko Albani (ESA) and Iolanda Maggio (ESA).

DATA-2: Full representation of CEOS Agency datasets in the IDN and accessibility via supported WGISS systems and standards – Q2 2019. Michael Morahan (NASA).

DATA-8: Maintain and evolve WGISS Connected Data Assets Infrastructure and Systems for CEOS agencies data and services discovery and access – Q4 2018. Yonsook Enloe (NASA).

DATA-9: ECVs/CDRs Discovery and Access through WGISS Systems – Q3 2018. Michael Morahan (NASA).

DATA-11: Data and Technology Exploration webinars and workshops - Q4 2019. Chris Lynnes (NASA).

DATA-12: CEOS data holdings reported in GEO – Q3 2018. Yonsook Enloe (NASA).

DATA-13: Review and update CEOS WGISS Best Practices to address data via user-defined polygons – Q1 2019; discussion about what this action is. Rob said this may be a matter of language and getting these reworded. To be removed.

DATA-14: Develop a White Paper on Single Sign-On (SSO) authentication – Q4 2018. ESA?

The following comments were made:

FDA: Robert Woodcock (NASA) inquired as to the activity of the ad-hoc team; Steven Hosford (CEO) replied that there have not been frequent meetings, but that many groups are actively working. The ad-hoc team meeting on April 17 will consolidate what teams have done in the last few months.

Data-2: Yonsook Enloe (NASA) commented that this has been an ongoing action that is very important and agency support is needed for this. Same with the ECVs – data providers have to register their data.

Data-9: Albrecht von Bargen (DLR) and Mirko said that WGISS analysing these based on the list form WGClimate. Mirko added that the motivation to do this lies in the benefits that the agencies will receive. Albrecht said these actions must be clear, with clear dates and communication to the agencies. WGISS already does the support.

Mirko also reported that Gilberto Camara (INPE) will be the next GEO Secretariat Director. The WGISS SLT representatives are participating in “GEOSS Platform” teleconferences. An ad-hoc meeting was held in March to address IDN visibility in the GEO-Portal. A CEOS-GEO bilateral meeting was held in January. The main outcomes were:

* GEO focusing on three Priorities of Paris Climate Agreement, Sendai Disaster Risk Reduction, and U.N Sustainable Development Goals. CEOS is supporting all three Priorities, particularly through leadership and contributions in the GEO-DARMA, GEOGLOWS, Blue Planet, EO4SDGs, etc.
* 2017-2019 GEO Work Programme will be released soon: 4 GEO Flagships, 24 GEO Initiatives, 10 Foundational Tasks and 35 Community Activities: CEOS contributing/leading several of them.
* CEOS co-organizing with GEO-SEC a session at the 3rd GEOSS Data Providers Workshop focused on establishing improved “User Metrics.” Currently planned for 3rd of May. Moderator: Mirko Albani.

Mirko addressed the following points for WGISS cooperation with CGMS

* WGISS best aligns with CGMS Working Group (WG) IV Global Data Dissemination. Many members of WGISS (e.g. NOAA, NASA EUMETSAT) are also members of CGMS.
* In the past, WGISS members have participated in WG IV to respond to CGMS action items. As an example, in response to CGMS action item # A44.03 (identify how far WGISS Interoperable standards were adopted), WGISS provided a list of CEOS agencies who have implemented CEOS OpenSearch Best Practices.
* CGMS-46 Working Group (WG) IV meeting will be held on 3-8 June 2018 in Bengaluru, India. Opportunity to present on behalf of WGISS a paper on the CEOS metadata model (e.g. DIF-10, etc.): NASA is sending participants, Andy Mitchell is trying to identify someone who could present on behalf of WGISS.

Mirko listed the following opportunities with other CEOS groups:

WGDisasters:

* Recovery Observatory (RO): Progress in Haiti with growing stakeholder engagement and positive feedback. The current RO Work Plan will continue with a three-day user workshop in May. New partners are also joining, including NOAA and EC. Opportunities for broader partnerships and extension to other countries in the region are under review. Discussions on a global, ‘generic’ RO are planned.
* Disaster Risk Management (DRM) Demonstrators: WGDisasters has identified potential linkages across CEOS WGs (WGCapD, WGISS) as well as GEO-DARMA. Three new Demonstrators are proposed to build on the Pilots (Seismic Hazards, Volcanoes, and Floods) and have been formulated and reviewed by the WG. The Seismic Hazard and Volcano Demonstrator teams are summarising their objectives, activities, benefit to users, and data needs, for consideration at SIT-33.

WGCapD:

* WGCapD’s Work Plan for 2018-2020 is under revision and will be completed by April.
* The WGCapD Secretariat will represent WGCapD at the WGISS and WGCV meetings at INPE.
* WGISS Technology webinars publicized through WGCapD. CEOS Work Plan Action: Promote awareness on FDA. Promote results of WGISS FDA activities through WGCapD. Define activities and contributions from WGISS members, then present through WGCapD from a user perspective. Opportunity: Store WGCapD training material / documentation on WGISS Web site. Opportunity: Joint discovery of data and training material associated with data. Need to define Metadata for knowledge associated with data. Hilcea Ferreira noted that WGCapD has many materials at the learning centre, but improvements are needed. Mirko will talk to Nancy at the SIT meeting regarding this.

CEOS Carbon Team:

* Carbon portal development ongoing. Prototype portal being updated based on 6 additional requirements received from Carbon team. Awaiting from Carbon Team (S. Plummer): Priorities for ECV/CDRs discovery and access. Priorities for In-situ data. Use case definition to drive development. Possible link to ICOS Carbon portal for in-situ data being evaluated. Opening for wider user community testing in July 2018 (to be confirmed)

WGCV: Addressed through Joint Session

WGClimate: ongoing through ECV inventory analysis

SEO: CEOS Work Plan Action FDA-5: Promote awareness on FDA. Opportunity: better address relation between COVE tool and WGISS Connected Data Assets Infrastructure. Opportunity: develop approach for on demand data-cube creation. Yonsook commented that COVE is harvesting the metadata and putting it in their database; they gained access within a day using the CEOS Open Search.

Virtual Constellations: will be addressed at SIT-33

Mirko reported that the following technology webinars have been coordinated by WGISS; he suggested enlarging this to the GEO communities and identifying topics that are more user-focused, like FDA and connected data assets:

1. Relevancy Ranking of Data Search Results
2. Data Cubes for Large Scale Data Analytics
3. Burgeoning Role of Python for Earth Observation Data Analysis
4. The OGC Coverage Standards Suite:  Introduction and Overview
5. Scrum and scale agile methods

Mirko gave the following note on WISP support: NOAA can no longer support the WGISS Infrastructure Services Project (WISP) which provides technological support and core services for WGISS activities (teleconference and web conference support via GoToMeeting, website support, mailing list and meeting archive support).

**Action WGISS-45-1:** Call for interest from a WGISS-participating agency to take over WISP: deadline for expression of interest is end of April 2018. WGISS Chair will ask SEO to support these services in case no other agency actively involved in WGISS volunteers to take over WISP.

Mirko concluded his report noting that WGISS needs to start thinking about a vice-chair for October 2019-October 2021.

## WGISS 2018-2020 Work Plan

Mirko Albani presented the draft outline for the WGISS 2018-20 Work Plan. Input for CEOS 2018-2020 Work Plan was reviewed with the WGSS-Exec and provided to the CEO with updates and revisions. The CEOS 2018-2020 Work Plan is currently being endorsed.

A WGISS 2018-2020 Work Plan is being drafted to cover routine activities and address the CEOS Work Plan actions on WGISS. A draft was circulated to the WGISS-Exec for comments/inputs. Finalization is expected by end April 2018.

Mirko displayed the organizational structure and table of contents for the work plan. Each action has an objective, deliverables list, projected completion date, background information, responsible CEOS entity, action owner and contributors, and additional comments. It is the aim of WGISS to participate in many CEOS activities rather than working independently.

A table of the WGISS horizontal and joint activities follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Action Objective** | **Deliverables** | **Projected Completion Date** | **Background Information** | **Working Group (IG)** | **Responsible** | **Contributors** |
| **WGISS-01:** Support for outreach activities | Production and distribution of outreach material (Brochures, Flyers, Video etc.) | Ongoing | WGISS recognises the need to promote the output of its work within its own agencies and external organisations in order to ensure the adoption of results and recommendations; and to ensure continued funding WGISS activities. | WGISS | Mirko Albani | All Interest Groups when requested |
| **WGISS-02:** Support for Conferences, Workshops and seminars | Production and presentation of papers at conferences, seminars and workshops | Ongoing | WGISS recognizes the need to sponsor and publicize its results and work outside taking part to conferences, workshops and seminars. | WGISS | Mirko Albani | All Interest Groups when requested |
| **WGISS-03:** Foster new liaisons for information and knowledge sharing | New collaborations opportunities | Ongoing | WGISS recognizes the need to share the information and knowledge of its work within its own agencies and external organisations in order to ensure the adoption of results and recommendations; | WGISS | Mirko Albani | All Interest Groups when requested |
| **CB-13**: Develop and/or deliver webinars for users in developing countries | Technical webinars | Ongoing | Webinars will provide information and training on advanced satellite Earth observation topics, such as data access/availability, data processing, and more. | WGCapD with the support of WGISS |  |  |
| **WGCV-1**: Data Formats and Interoperability in the frame of FDA |  |  |  | WGISS/  WGCV |  |  |
| **WGCV-2**: Quality Indicators in Discovery Metadata |  |  |  | WGISS/  WGCV |  |  |
| **WGCV-3:** CEOS Data Cubes and CEOS Test Sites Data Access in support to WGCV Activities |  |  |  | WGISS/  WGCV |  |  |
| **WGCV-4**: Standardization and Best Practices |  |  |  | WGISS/  WGCV |  |  |

Preservation and Stewardship:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Action Objective** | **Deliverables** | **Projected Completion Date** | **Background Information** | **Working Group (IG)** | **Responsible** | **Contributors** |
| **WGISS-04:** Drafting of Best Practices and guidelines | Drafting of new Guidelines, Best Practices, White Papers and Recommendations | Ongoing | WGISS exploits knowledge of its participating agencies for the benefit of all CEOS and non-CEOS members by producing recommendations and guidelines concerned with all aspects of data and information management. | WGISS (DSIG) | Mirko Albani | DSIG |
| **WGISS-05:** Organize topic sessions on data stewardship to exchange information and lessons learned | Presentations and related material | Ongoing | Data Preservation and Stewardship sessions on topics of interest for CEOS agencies. | WGISS (DSIG) | Mirko Albani | DSIG |

Discovery and Access:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Action Objective** | **Deliverables** | **Projected Completion Date** | **Background Information** | **Working Group**  **(IG)** | **Responsible** | **Contributors** |
| **WGISS-06:** IDN Guide | Procedure and Guide Document | Q4 2018 | How-To Guides to tag datasets in the IDN for granule search via CEOS OpenSearch | WGISS | Michael Morahan |  |
| **WGISS-07:** FedEO Guide | Procedure and Guide Document | Ongoing | Review and update FedEO How-To Guides for Data Partner and Client Partner | WGISS | Andrea Della Vecchia |  |
| **WGISS-08:**  CWIC Guide | Procedure and Guide Document | Ongoing | Review and update CWIC How-To Guides for Data Partner and Client Partner | WGISS | Yonsook Enloe |  |
| **WGISS-09:** WGISS Connected Data Assets Guide | Procedure and Guide Document | Q4 2018 | Produce a How-to Guide for Client partners that describes how to connect and search for data in IDN, FedEO, and CWIC. | WGISS | Yonsook Enloe | System Level Team |
| **CARB-15:** Carbon data  Portal prototype | Deployment of a Carbon Data Portal Prototype | Q3 2018 | Implement a carbon data portal to facilitate the discoverability and accessibility of ECV products and space-borne CDRs. The portal is designed with a service-oriented architecture and follows the principles outlined by the GEOSS Community Portal white paper. The portal will seamlessly access data both in CWIC and FedEO to provide necessary data, tools and services to the carbon science community of both CEOS and GEOSS. The reference implementation can be shared with the broader CEOS carbon community. | WGISS | Ken McDonald,  Martin Yapur |  |
| **DATA-2**: Full representation of CEOS Agency datasets in the IDN and accessibility via supported WGISS systems and standards | Procedure for IDN registration and periodic reports on registered datasets | Q2 2019 | As the IDN contains OpenSearch endpoints for data access and is also the CEOS Data Collections access point for the GEOSS Common Infrastructure (GCI) and GEO Portal, it is essential that all CEOS Agencies keep information on their data collections, including Analysis Ready Data, up-to-date in the IDN according to its metadata model (DIF-10) | WGISS and CEOS agencies / (ALL) | Michael Morahan,  Mirko Albani | System Level Team |
| **DATA-8**: Maintain and evolve WGISS Connected Data Assets Infrastructure and Systems for CEOS agencies data and services discovery and access. | WGISS Connected Data Assets system description document and periodic reports describing the status. | Q4 2018 | Consolidation, operation, maintenance and evolution of current CWIC/FedEO/IDN overall Architecture | WGISS | Yonsook Enloe,  Michael Morahan,  Andrea Della Vecchia | System Level Team |
| **DATA-9**: ECVs/CDRs Discovery and Access through WGISS Systems | Periodic Reports on ECVs/CDRs registrations | Q3 2018 | Facilitate discoverability and accessibility of ECV Products and space-born CDRs relevant for the CEOS Carbon Action via WGISS Connected Data Assets Systems & Standards (FedEO/CWIC/IDN, OpenSearch). | WGISS | Michael Morahan,  Andrea Della Vecchia,  Iolanda Maggio | System Level Team |
| **DATA-12**: CEOS data holdings reported in GEO | Production Periodic Reports | Q3 2018 | Provide support to GEO in their efforts of reconciling metrics of CEOS data holdings provided through WGISS Connected Data Assets standards and systems. | WGISS (SLT) | Yonsook Enloe | System Level Team |
| **WGISS-10:** Production of video on WGISS Connected Assets | Delivery of a video | Ongoing | WGISS Connected Assets activities publicity through short video. | WGISS (DSIG) | Yonsook Enloe | System Level Team |

Notes:

WGISS-10: Yonsook noted that the development of the previous video was a lot of effort. She recommended a more professional production. There could be one for Connected Data Assets and another for Data Preservation.

Data-12: Yonsook commented that their GEO’s metrics are not reported correctly since there is data duplication. WGISS has a page with metrics that are constantly maintained. Steven Hosford asked for more coaching on this so that he can properly communicate it. Mirko suggested interacting with the GCI, who harvests granules from the agencies, from CWIC, and from FedEO. It is important that what they make discoverable makes sense and that their metrics are accurate.

Interoperability and Use:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Action Objective** | | **Deliverables** | | **Projected Completion Date** | | **Background Information** | **Working**  **Group** | **Responsible** | **Contributors** | |
| **FDA-5:** Promote awareness of FDAs. | | Material for FDA promotion | | Q4 2018 (Interim reports) | | With growing interest in Future Data Architectures, WGCapD will identify required CB material to support rollout of FDA technologies. Materials will range from basic introductions to information relevant to decision makers. | WGCapD supported by WGISS |  |  | |
| **FDA-8**: Establish a common description of Future Data Architecture functional blocks and identify interfaces and interoperability approaches. | | FDA Functional Building Blocks description with interfaces | | Q3 2018 | | Based on the outputs of the inventory and review of existing standards and approaches at CEOS agencies, on the pilot projects and using the various workshops (listed below) where FDA activities are discussed, FDA-AHT will establish a common understanding and develop a white paper describing the functional blocks and typical interoperability approaches for a generic FDA. | WGISS |  |  | |
| **FDA-9**: Inventory and characterise existing FDAs operated by both public and private entities including the standards and approaches they use (e.g. Data Cubes, Exploitation Platforms, Copernicus DIAS, etc.). | | FDAs inventory | | Q3 2018 | | As CEOS agencies are defining their processing and data dissemination standards, they seek to apply and follow international standards and best practices, including those generated by WGISS.  This does not only concern common standards in terms of catalogues, metadata, terminology, and semantics, but it also involves interoperability standards for data discovery and download and for EO data analysis Application Programming Interfaces (APIs), as well as common interface standards such as INSPIRE, OGC, and W3C, and interoperability with other data access services (e.g., European Data Portal, international, GEOSS). WGISS will inventory and review existing standards and approaches in use by different FDA systems and platforms at CEOS agencies (e.g. Data Cubes, Exploitation Platforms, Copernicus DIAS, etc.). | WGISS |  |  | |
| **FDA-10**: Finalise inventory of Software and Tools available or used at CEOS agencies for EO data exploitation and use focusing on Open Source but remaining as broad & inclusive as possible and implement a mechanism for discovery and access. | | Inventory of SW and Tools | | Q3 2018 | | Each CEOS agency will continue to develop its data and computational infrastructures consistent with its capacity and user service mandates. CEOS has a role in identifying tools to support complementarity and interoperability across CEOS agencies in support of the FDA strategy objectives. WGISS will finalise the ongoing work of inventorying the software and tools available or used at CEOS agencies for EO data exploitation and use (e.g. EO data visualization, analysis, processing, readers/writers, etc.), and implement a mechanism for discovery and access. Focus will be on Open Source but remaining as broad and inclusive as possible. | | FDA AHT supported by WGISS |  |  |
| **FDA-11**: Organize several sessions/workshops to share lessons learned and outcomes from FDA systems and platform pilots and Interoperability Projects. | | Workshop/Sessions for sharing of lessons learned and outcomes | | Throughout 2018 until Q4 | | Several CEOS agencies have already carried out initial pilot projects and CEOS seeks to share the experience gained from these to ensure the discussions on the strategic dimensions of FDA are informed by practical evidence. Sessions (including but not limited to those detailed below) should focus on several issues ranging from technical to programmatic lessons learned will be shared and considered as input for the definition of technical FDA recommendations. | | WGISS |  |  |
| **FDA-13**: Develop a User Metrics Best Practices. | | User Metrics Best Practices | | Q4 2018 | | The proposed User Metrics initiative seeks to ensure planning and responsibilities are put in place for CEOS to leverage the experience being gained by individual agencies and to have an ongoing effort to collate available user metrics and to adapt the FDA strategy as these metrics are analysed beyond what is simply being downloaded. WGISS will perform a survey on existing user metrics in Earth Observation and other domains (e.g. social media) and develop a best practice for User Metrics recommended for application by CEOS agencies; these will include a user questionnaire to allow classifying the final user in terms of use of the data products and allowing generation of summary statistics on how EO data is being used and for what. | | WGISS |  |  |
| **DATA-13**: Review and update CEOS WGISS Best Practices to address data via user-defined polygons. | | Updated standards | | Q1 2019 | | Existing standards and best practices in use for CEOS data access will be revised to address “Clip and Ship” access based on user-defined polygons. | | WGISS |  |  |
| **DATA-14**: Develop a White Paper on Single Sign-On (SSO) authentication. | | White Paper | | Q4 2018 | | Single sign-on (SSO) allows user login with a single ID and password to gain access to connected (federated) systems. This capability is crucial for interoperability between different FDA platforms and systems. WGISS will develop a white paper on single-sign-on (SSO) authentication best practices to support machine- to-machine authentication for EO analysis services. | |  |  |  |

Notes:

FDA-5 and FDA-8: These need to close quickly. Steven said these actions involve characterizing the EO ecosystem and WGISS is involved in terms of interoperability; other aspects have already been started. There remains a challenge with terminology; Steven will discuss some of this. He added that the conversations on leadership have been happening but there needs to be an ad-hoc team meeting. Rob commented that the completion dates shown may be too soon.

FDA-10: It is unclear who really is responsible for this. Hilcea suggested that it would be helpful to make tools and training materials discoverable. Steven said this is a better fit for WGISS because SEO is focused on the CEOS Data Cube.

FDA-11: This could be linked to the technology exploration webinars but Steven said it is more about workshops than about webinars.

FDA-13: Input from Andy Mitchell is needed.

Data-14: Esther Conway (UKSA) said that UKSA is interested to know what others are doing in the area of security, and Mirko requested that contribution UKSA contribute to this.

Technology Exploration

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Action Objective** | **Deliverables** | **Projected Completion Date** | **Background Information** | **Working Group**  **(IG)** | **Responsible** | **Contributors** | **Notes** |
| **DATA-11**: Data and Technology Exploration webinars and workshops | Technical Webinars | Q4 2019 | WGISS will host at least one workshop annually to serve as a forum for exchange of technical information and lessons-learned experience about current, trending and future data management approaches and technologies, services and other Internet-related technologies. | WGISS (Technology Exploration Interest Group) | Chris Lynnes | Experts team |  |

ESA will update the draft work plan to include the results from WGISS-45, and will assign sections of the plan to individuals for completion. The final draft for review with WGISS-Exec is planned for end of April and formal issue by May.

**Action WGISS-45-2:** ESA to update the draft WGISS Work Plan with the results of WGISS-45 session discussion; final draft to WGISS-Exec for review by end of April; issue by end May 2018.

**Action WGISS-45-3:** Michelle Piepgrass to organize a telecon in May (WGISS, WGCapD, SEO, CEO) to identify ways for WGISS to support WGCapD on CEOS Workplan action FDA-5.

## WGISS Brochure

Yonsook Enloe (NASA) gave a presentation of the WGISS notion for creating one-page flyers. These would be developed by an interest group or project, targeted to a specific audience with a simple message to achieve a purpose, and using easy-to-understand terminology. The soft copy could be stored on the WGISS website for easy distribution to WGISS representatives and to allow making local copies.

Yonsook displayed a sample flyer targeted to those who develop software tools, using three different graphical backgrounds. She noted that Andy Bingham and Martin Yapur assisted with the development of the flyer.

It was agreed that this is a good way to get a targeted message across, especially when attending meetings and conferences. There was discussion on the specific content of the flyer presented, including the landing page that the reader is directed to. Rob suggested that the landing page include the option to actually attempt the data access, and explanation on to motivation to go to WGISS instead of going to the agencies’ data hubs.

Yonsook commented that this flyer is a message for the client partners. The idea is to provoke interest, and to provide a place where they can ask for information.

Next steps are to select a graphic background, place the flyer on the WGISS website, and encourage participants to distribute it.

Regarding the full brochure, Mirko suggested reviewing and finalizing the soft copy.

**Action WGISS-45-4:** Michelle Piepgrass to provide Interest Group leads the WGISS brochure (soft copy) presented at WGISS-44; Interest Group leads to review and simplify their text, keeping the mission of WGISS by April 25. WGISS Exec to consolidate and finalize the brochure, and create a page on the WGISS website for easy access.

# Data Preservation and Stewardship

Mirko Albani introduced the Data Stewardship Interest Group (DSIG) session and gave an overview of their activities. He described their collaboration with ESIP, with ISO TC 211, and the WGISS Maturity Matrix. One action for WGISS is to set up a web page on Lessons Learned and Best Practice documents and it was suggested to take this to the CEOS level, covering all the working groups. This can be placed at the “Our Work” level of the CEOS website. Mirko gave an example of what the page would look like.

Mirko also discussed an action to put together a procedure for making new datasets visible to CEOS. He suggested placing it under the WGISS node of the CEOS website, and on the Discovery and Access page. The main CEOS web page could also contain a flashing “News” bulletin that links to the relevant page. This will need a responsible entity to continuously monitor the availability of new datasets. Michael Morahan mentioned that the IDN has a “what’s new” section for newly registered datasets in the IDN. The message to CEOS is that all datasets should be registered in the IDN.

Mirko listed several recent and upcoming conferences related to data stewardship, and provided links to these.

## DSIG Standards Status and Way Forward

Iolanda Maggio gave a presentation on DSIG standards status and way forward. She listed the documents released by the group, which are also displayed in the document management table on the WGISS website.

New opportunities and tasks include:

* Create a White Paper on Data Stewardship Reference Model (not more than two pages) and Maturity Matrix to send to the DMP team as a proposal for inclusion in the GEO DMP.
* Appraisal procedure: A structured appraisal of a dataset considered for long-term preservation helps assess if an Earth Observation dataset is valuable enough to be preserved for future generations (e.g. high level). It also provides an estimate of the cost and effort involved, and of any associated risks.
* Update Data Preservation Guidelines and Preservation Workflow vs. DMP, WGISS Data Management and Stewardship Maturity Matrix and Common Framework for Earth-Observation Data.

Iolanda described the intended audience and themes of the proposed white paper.

Mirko suggested that the white paper appraisal could wait, but that the initial paper would be helpful. Activities at the DMP level are not frozen and WGISS should have something concrete to propose.

Greg Stensaas (USGS) commented that John Faundeen (USGS) is doing a lot of this work and it would be good to obtain his participation or request that he review.

Iolanda mentioned that a lot of feedback was received for the Maturity Matrix (MM). Richard reminded that there have been a few MMs developed and that Ge Peng is very satisfied with WGISS’ because it includes input from NOAA, and it is aligned with the DMP.

## Data Management and Stewardship Maturity Matrix Lessons Learned

Iolanda Maggio gave updates on the WGISS EO Data Management and Stewardship Maturity Matrix (MM) as applied to the AVHRR Reprocessing Project. She described the project’s requirements, measurement, gap evaluations, and work logic and strategy. They found that using the MM, one can have clear results in terms of level in the rows of the matrix and of areas in the columns of the matrix. It is easy to see which activities need to be done, and apply this to the vision of the agency in order to compare it to available resources.

Iolanda reported that the exercise was extremely useful; they found that it is a way to evaluate the data stewardship processes and to plan the goals of the data stewardship processes and projects.

Greg Stensaas commented that one of the challenges is finding the common definitions for interoperability and traceability, and it would be helpful to include this in the white paper. Iolanda referenced the glossary that WGISS has prepared, and said that a section on definitions can be added.

Way forward for the team is collection of Lessons Learned and analysis and tuning of the White Paper, in collaboration with NASA/ESDIS as joint activities.

The team is also awaiting possible input from the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix (DSMM) Group managed by Ge Peng.

Iolanda also noted the publication in the Journal of Global Change Data and Discovery by Liu Chuang (Institute of Geographic Sciences and Natural Resources Research).

**Action WGISS-45-5:** Mirko Albani and Iolanda Maggio to organize development of a procedure for making visible to CEOS the new agency datasets that become available by Q4 2018.

**Action WGISS-45-6:** DSIG to create a White Paper on the Data Stewardship Reference Model and Maturity Matrix and send it to the DMP team as a proposal for inclusion in the GEO DMP by July 2018.

# Data Discovery and Access

Yonsook Enloe introduced the data discovery and access session.

## WGISS Connected Data Assets

Yonsook Enloe (NASA) reported that the WGISS Connected Data Assets includes access from the International Directory Network (IDN), FedEO (Federated Earth Observation Gateway), and CWIC (CEOS WGISS Integrated Catalog). WGISS-supported standards are now used to search and access data from thousands of collections and hundreds of millions of inventory records, with additional data collections and inventory records from CEOS agencies being added daily to this integrated system.

The WGISS Connected Data Assets also provide access to CEOS agency datasets through their integration with the GEOSS Common Infrastructure (GCI). More and more, CEOS agencies continue to adopt the WGISS-supported standards and make their data discoverable via the WGISS Connected Data Assets.

The WGISS Connected Data Assets System Level Team (SLT) handles coordination and oversight of the operations and evolution of this integrated system and also provides technical support for CEOS partners that offer access to data, and to client partners that connect to the data sources.

Yonsook gave the following up-to-date metrics, which are displayed on the WGISS website:

* Discoverable collections (via IDN): over 32,000+ (over 15,000+ GEOSS DataCore)
* Searchable granules: over 296+ million (more added daily from live missions)
* Contributing agencies: ESA, NASA, NOAA, USGS, ISRO, INPE, EUMETSAT, CNES, etc.
* Additional agencies working on connecting data assets: ChinaGEOSS, China (NRSCC), NOAA One-Stop

The team needs to reach out to more potential tool developers. Ideas are to work with SEO and CEO to distribute the one page flyer described earlier, and for WGISS representatives to do the same.

Yonsook listed additional work to meet the WGISS Work Plan deliverables:

* How-to guides to tag datasets in the IDN for granule search via CEOS OpenSearch
* Review and update FedEO how-to Guides for Data Partner and Client Partner
* Review and update CWIC how-to Guides for Data Partner and Client Partner
* Produce how-to Guides to become WGISS Client Partner that describes how to connect and search for data in the IDN, FedEO, and CWIC

In regard to GEOSS, integration continues, and the IDN is now registered in the GEOSS Yellow Pages. Email and telecon discussion about adding “IDN” to the GEO Web Portal 🡪 “GCMD/IDN” is occurring. The FedEO and CWIC are accessible from GEO web portal and integrated via GEO-DAB, and the team maintains participation in the GEO Data Providers workshop. The WGISS Connected Data Assets also keeps the webpage on data metrics up-to-date, and has given the information to the GEO Sec.

Working toward a WGISS federation involves data collection registration at the IDN using the IDN keywords; information about how granule search is supported will be included in the data collection registration with tags for granule access. Data partners need to support one of the two supported WGISS standards (CEOS OpenSearch Best Practices (v 2) or OGC CSW 2.0.2)

All searchable data must have a data access path, servers must have high availability, and a technical POC is needed for each data partner.

Recent and future happenings include:

* Agreement on the resolution of the interoperability issues and documented those in the CEOS OpenSearch Best Practices document
* Team review of OGC 17-003 (OpenSearch-EO GeoJSON Response Encoding Standard) completed
* Team review of OGC 13-026r9 (OpenSearch Extension for EO) completed
* FedEO data collections being registered in the IDN; datasets registered in the IDN will be tagged, and FedEO OS access to FedEO datasets in IDN will be directed to FedEO
* ECV collections will be registered in the IDN (Mirko contacting data providers)
* Continuing integration with GEOSS
* CWIC Smart Client Validator for CEOS OS conformance testing
* Conformance Test document led by ESA
* Will be producing WGISS Client Partner Guide

Makoto asked about JAXA being in the WGISS Assets Metrics. Yonsook replied that they need to make their client accessible.

Greg asked how data can be made accessible. Yonsook replied that all that is required is an access point, with IDN registration, granule search, and unrestricted data.

Andrea noted that there are two different standards:

* OGC 17-047: OpenSearch GeoJSON(-LD) Response Encoding (Open Search returned metadata encoded with GeoJSON(-LD))
* OGC 17-003: EO Product Metadata GeoJSON(-LD) Encoding (Granule Metadata model)

Esther asked about UKSA’s data; Michael replied that data going to FedEO can be pushed to the IDN.

Chris commented that NASA has endorsed OpenSearch as a technical note.

The WGISS Best Practice document needs to be added to the WGISS documents page.

**Action WGISS-45-7:** Yonsook Enloe, Richard Moreno, Rob Woodcock, Mirko Albani, and Chris Lynnes to identify and submit to Michelle the valid versions of their Best Practice and White Papers. Michelle to display these on the Document Management Table on the WGISS website by end of June 2018.

**Action WGISS-45-8:** Mirko Albani, Iolanda Maggio and Michelle Piepgrass to address Best Practice and White Papers (not lessons learned) Web page (to be put at highest level in CEOS Web Site) population with other WGs material after completion of action on WGISS Interest Groups. Maintain and populate page by September 2018.

## International Directory Network (IDN) Report

Michael Morahan gave the International Directory Network (IDN) report. He discussed the transition from DIF-9 to DIF-10, listing the new and restructured fields, the benefits and reasoning for the transition, and the transition schedule and method. Michael also discussed the GCMD keyword updates, the DIF-10.3 updates, and the UMM-C 1.10 updates. Michael described methods for validation of metadata records, and the docBuilder tool for the new DIF-10.

Michael concluded with a discussion of the mapping of the WGClimate ECV Inventory. The team took the repository with the goal to identify the ECV records within the IDN. He reported that of the 913 ECV records, 496 are current records and 417 are future records. Of the 496 current ones, 288 were found to be IDN granule/product accessible via CWIC infrastructure, four from the CNES infrastructure, and 204 were not in the IDN.

ESA sent emails of inquiry for the 204 not in the IDN. Seven (of 25) replies were received, containing request for more details on the registration process and infrastructure, and information about obsolescence of the requested datasets

Mirko commented that WGISS needs to provide the ECV feedback to WGClimate. Yonsook noted that this a great example of cooperation between working groups, and it should be highlighted at the SIT level.

Steven suggested sending a reminder to those who have not responded, and that Mirko report it at the SIT.

Rob asked how these will be tagged; the tags will be given to WGClimate.

## FedEO Status

Andrea Della Vecchia gave a presentation on the FedEO Status. He began with an architecture overview and continued with recent activities. Andrea gave details on various aspects of the client – the landing page, the basic and advanced search, the TABs, and the product search example. He also discussed the gateway:

* Support for virtual dataset series (defined at OSGW configuration level)
* OSGW now supports the OpenSearch metalink 3.0 and 4.0 media types for dataset metadata discovery
* OSGW provides an OpenSearch GeoJSON interface for dataset series metadata and dataset metadata discovery [OGC 17-047] & [OGC 17-003]
* OSGW supports the “strict” parameter to return exceptions in case a submitted search parameter is not supported by server
* OSGW returns an “up” atom:link in case of dataset search allowing to retrieve the corresponding dataset series metadata
* Improvement of the OSGW OSDD with configurable title attribute to provide explanation on some search parameters, explanation that could be displayed by a client to help the user

Andrea reported that the collection catalogue content has been updated with the goal to make the collections available to the IDN. There is also a new metadata editor and product catalogue. He also listed metrics of collections accessible through FedEO.

Future activities include and ESA catalogue with interfaces that are publicly available with OGC OpenSearch specifications, or CEOS best practice. They also plan to work in FedEO scalability technology.

In response to a question about data duplication, Yonsook said that the user decides in the first search which collection to use when two of the same data are available. Lubia brought up this problem because it means that the agency has to repeat the registration of the same data that belongs to different collections. Rob suggested that WGISS address this topic in the near future since it is only getting worse and it affects the user experience. However, the data may or may not be there, and the correct registration in the IDN is important. Michael reminded that this is something that was dealt with when registering GHRIST, and it is the reason that the metadata fields for the archiver and the data provider were added.

Andrea concluded with a listing of the links for FedEO access.

## CWIC Status

Yonsook Enloe gave a status on CWIC. She reported that they are continuing to support the WGISS System Level Team, providing technical support for CWIC connectors, support for GEOSS, and assisting in the production of the client partner guide.

She gave a brief status on each of the CWIC partners, and the requirements data partners, and the resources available to them.

## China (NRSCC) and China GEO Status

Eugene Yu gave a presentation on the data discovery and access status of the Chinese data. He noted that the CWIC team is working with two data providers in China: ChinaGEOSS and China NRSCC.

ChinaGEOSS is a national satellite remote sensing data facility of National Remote Sensing Center of China (NRSCC), under the Ministry of Science and Technology, China. He gave details of the satellites and sensors, and gave metrics on granules of these missions. In terms of the CWIC connector development, the first bi-lateral teleconference of technical teams occurred in November with continued technical exchanges. NRSCC has requested to sign an MOU between NRSCC and CEOS for ChinaGEOSS to join CWIC.

China NRSCC is funded to produce global annual land ecological products freely available to worldwide users. It includes multiple data sources and algorithms, and long-term data products. More than 30727 granules are registered in the NRSCC CSW catalogue.

## Carbon Portal and Strategy and Demonstration

Kenneth McDonald (NOAA) gave an update on the Carbon Portal and Strategy, with a demonstration by Liping Di. He listed the objectives of the portal and described the collection-level search and the granule-level search. Liping demonstrated the spatial and temporal constraints of the portal.

Ken explained that the search and access can be done with predefined carbon regions and predefined carbon topics.

The portal will be presented at the SIT as a cooperation with GEO.

Mirko asked if the six additional “needs” sent previously by the Carbon Team have been incorporated. Ken replied that most of them are there. Mirko asked if WGISS should ask (again) the Carbon team to provide use cases from the Carbon community; this is important.

**Action WGISS-45-9:** Liping Di to send an email to Mirko Albani with the details of information he still needs from the Carbon team. Mirko to forward these to the CEOS Carbon Team and organize a teleconference to review the status of the portal so that it can be opened to the broader community for review by April 20, 2018.

## OGC Documents Specification Status

Andrea Della Vecchia gave a report on OGC Documents Specification for OGC 17-003 and OGC 13-026r9. He began with an overview of activities and discussed:

* OGC 13-026r9: EO Extension for OpenSearch
* OGC 17-047: OpenSearch GeoJSON(-LD) Response Encoding
* OGC 17-003: EO Product Metadata GeoJSON(-LD) Encoding

Andrea’s conclusions are that the SWG addressed the identified weaknesses:

* Original specification OGC 13-026r8 incomplete, lack of coherence -> extensive “CEOS Best Practice” needed.
* “Poor interoperability of search results (different metadata models)”
* “Atom/XML not well suited for Web client consumption”
* “Need for linking EO resources to services “bring user to data paradigm” (Exploitation Platforms)”

Next steps include:

* 2/2017: EOP SWG reactivated, initiated documents (13-026r9, 17-003)
* 3/2017: TC Delft (presentation of rationale, targets, first results...)
* 4/2017-9/2017: Work done, split off GeoJSON(-LD) encoding (17-047)
* 09/2017: TC Southampton, status presentation, discussion of CRs
* 09/2017-01/2018: Fixed all CRs, ATS, examples, other improvements…
* 02/2018: SWG votes to release candidates standards for OAB review
* 03/2018: in case of acceptance OAB may give permission for public comment
* 03/2017-04/2018: 30-day public comment period
* 03/2018: OGC TC Orleans: SWG summary presentation of specs in Closing Plenary (10 min), Request for e-vote (during public comment, so TC gives permission to start vote once public comment ends and public comments (if any) are addressed)
* 05/2018: SWG to modify candidate standard based on comments and resubmit
* 06/2018: TC adoption vote 45 days, and 2 weeks day IPR review period.

Future work:

* First draft of OGC 17-084: GeoJSON(-LD) Encoding for EO Collection Metadata
* Complete migration from XML to JSON

FedEO has to comply with INSPIRE; therefore in terms of metadata model all the INSPIRE requirements should be included.

Chris asked if there are proofs-of-concept for returning the response on GeoJSON instead of Atom. Andrea said that this was discussed at WGIS-44. He added that it may not be mature enough to try it for next year’s OGC testbed.

## CEOS OpenSearch Project Report

Andrea Della Vecchia (ESA) gave a report on the CEOS OpenSearch Project. He began with a brief history of the OpenSearch project and document development. The current activity involves the development of a Conformance Test Plan Document. He displayed the proposed table of contents and noted that the first draft is expected to be complete by the end of April, and submitted to the WGISS SLT for review.

Short term plans include the consolidation of the CEOS Open Search Conformance Test document and the CEOS Conformance Test software. In the long term, a further review cycle of CEOS Best Practice, considering OGC 13-026r9 and OGC 17-047, would significantly simplify the document.

Andrea also discussed synergies with OGC initiatives.

## Data Asset Enhancement — Synergies with FDA

Robert (Rob) Woodcock (CSIRO) gave a presentation on data asset enhancement and synergies with FDA. He began with a review of current data cube and FDA activities

Rob stated that user requirements are fairly well known, with routine discovery, download, processing and ARD (or similar), and the need to hit multiple CEOS agency collections using machine-to-machine APIs. The user experience is quite variable.

User requirements are demonstrably met, but issues in the EO supply chain are multiplied in FDA environments.

An opportunity for WGISS is an FDA demonstrator that does the following:

1. Characterise the user experience in accessing CEOS agency data for new analysis architectures
2. Tidy terminology – as is done in Open Search Best Practice document
3. Gap analysis
4. Develop WGISS revised notional architecture (FDA-8, part of FDA-9)
5. Develop Recommendations for CEOS.

There are some outstanding thoughts/issues with FDA -10 inventory of tools and FDA-12 inventory of data product formats used.

Recommendations from Rob are that WGISS draft a Call for Participation in an FDA Demonstrator Proposal with focused terms of reference to gather missing information, and to supplements the existing workshops which report on FDA activities. A WGISS committed core team could provide collation of information and deliverables, and convert the information into coordinated actionable outcomes

Steven commented that this may be a very productive approach, and he would support working on that. He highlighted the need to get feedback from a broad sample of users.

Rob concluded saying that a draft call for participation with clear terms of reference and scope would be helpful.

.

# Data Interoperability and Use

## Recovery Observatory

Richard Moreno (CNES) gave a report on the Recovery Observatory (RO). He began with the objective and rationale of the project, and described the planning beginning with 2013.

The Recovery Observatory has been a good example of cooperation between two CEOS working groups (WG Disasters and WGISS). WG Disasters has participated in the design and the development of the Recovery Observatory. The schedule was tight, but Recovery Observatory was developed on time.

Richard described the main features, which include a web portal that allows for editorial content and collaborative groups.

The triggering of the RO was decided by CEOS Chair in consultation with CEOS principals on December 22, 2016, after Hurricane Matthew (October 2016). The second mission to Haiti 29 May – 2 June 2017 consisted of an RO users’ workshop and feedback on sample products. The third mission to Haiti 5 Dec - 8 Dec 2017 consisted of a technical review to work within each thematic sector and link with universities. Richard listed preliminary results, and mentioned a showcasing event in May.

Chris asked about data latency; Richard replied that the work is long-term, tracking recovery, so latency is not a concern.

Mirko asked if anything further is envisaged for WGISS with this project; Richard replied that the tool is working and the work is done. The RO is well adapted to an earthquake; further development may be needed for a different kind of event (like a volcano). WGDisasters is discussing extending this project to other countries. If they come up with new ideas, WGISS would be ready to participate if the effort is pertinent.

The RO can be put this on the Open Source website.

## Open Source Software

Iolanda Maggio gave a presentation on Open Source software (OSS). She began with an overview of type and typology of OSS, and noted ESA’s policy, its objectives and justifications, and key points. The affected domains are:

* EO tool boxes
* Multi-mission components of ground segments infrastructures
* EO data Processors
* Operational software in general
* Software developed in R&D
* Contribution to OGC, WGISS
* Earth Science Exploitation platform

WGISS plans to create an OSS webpage; it should be placed in the Interoperability and Use Area of the website. Next steps are an OSS survey of CEOS agencies and reporting the results with relevant metadata.

Mirko commented that the goal is to link to the software repositories. Yonsook commented that agencies are hesitant because technical support may be expected. She recommended clear messages the availability of technical support, and whether the owner would accept suggestions for changes. Mirko said that an alternative to technical support is participating in forums.

Lubia recommended GitHub, since it provides a community. She would avoid layer upon layer to find things and to simplify the access.

Esther asked about EC software, but it was felt that the EC software is too broad, and WGISS is only looking at CEOS agency software. However, a connection to EC would be a great resource.

Rob said that CSIRO has something a bit more complex than what is suggested; the user can also search by problems to solve and by citations. He will follow up with WGISS with information on their system. They already have an information model, so there may be no need to reinvent it.

Esther mentioned data analysis platforms, wondering how to achieve getting licenses and obtaining a collaborative approach. She recommended a discussion on how to handle data analysis platform licensing. A session could be organized for WGISS-46.

**Action WGISS-45-10:** Iolanda Maggio to obtain from Robert Woodcock the CSIRO information model for managing OSS. Iolanda to distribute revised software metadata model to agencies, asking them to complete the relevant information and return to Iolanda by mid-May.

## OGC Testbed 13 and 14

Cristiano Lopes gave a presentation on OGC Testbed (TB) 13 outcomes and information on the OGC Testbed 14. He began with a background of the project, and listed the outcomes (reports, and a video).

Cristiano described the OGC EO Exploitation Platform Hackathon coming in May, with goals to advertise the TB-13 findings and collect feedback. There are a lot of interesting interdependencies for TB14. Cristiano also described the TB14’s three threads and schedule highlights.

Mirko asked if anything is needed from WGISS or CEOS agencies. Cristiano replied that he would welcome WGISS’ review of the reports, and review of the different options. He added that anyone who wants to be an observer at OGC can do so; the meetings are on Mondays, 11:00 US Eastern time. Chris Lynnes indicated interest.

## User Metrics Introduction

Iolanda Maggio introduced the user metrics session. One motivation for this session is the CEOS Work Plan action to develop a User Metrics Best Practice.

Iolanda recommended a survey on existing user metrics in Earth Observation and other domains, followed by development of a best practice for application by CEOS agencies. The BP would contain recommended metrics for user and data monitoring, “Satisfaction” questionnaires/survey, and user and data statistics. She mentioned a few points on what these metrics consist of, and when and how this could be done.

Iolanda provided the following points of discussion:

* Definition of “active users” - A user registered could be defined “active user” or perhaps it is better to measure actions like download, results of PI project, etc.
* Metrics availability - Should the metrics be available to everybody or should data usage figures be available only to the data provider supplying the data.
* Access to metrics - should an anonymous user have access to the metrics (full/partial)?
* Open data will increase the difficulty to collect metrics using only user logins or agency portal access. Recommendations are needed.
* Once identified, decisions are needed on analysis tools, discovery methods, and measurements and statistics.
* The commercial/business environment and markets could be really useful to obtain suggestions and processes because this topic started in that context.

Discussion:

* User satisfaction: what type, how to measure? Self-registration, survey, questionnaire, “like”. Big symposium (2-3000 people), doing surveys with promotions to encourage participation. Audience interaction made easy (sli.do), privacy, and security. Some combinations may become sensitive.
* Who are the users: commercial, scientific?
* What are they using: raw data, complex systems?
* Which metrics: having a commonly accepted type of metrics will simplify the issue. Usage and domain is not enough, have to move farther.
* Why do this: Use the metrics to understand the value of the data.
* What are users doing with the data: asking them to adopt DOI to see how the data is being referenced? This is deeper than user profiles. It will be very interesting to see how far you can push that without impeding on privacy.

WGISS will need to address the user metrics BP, and begin putting it together.

**Action WGISS-45-11:** Mirko Albani and Iolanda Maggio to ask every agency to send information on how they collect user metrics, which metrics they collect, and what they do with their metrics by end of April.

### ESA User Metrics

Mirko Albani gave a presentation of user metrics at ESA. He began with examples of metrics from a variety of ESA missions, thematic exploitation platforms, and projects. Mirko described metrics collected from users accessing data hubs, listing metrics on access, numbers, user geographical distribution, user types of activity, and scientific publications. He also described the Living Planet Symposium Survey Statistics.

In response to a question from Makoto, Mirko said that they ask users to credit ESA for using the data, but there is no way to enforce this. They are beginning to use DOI to add references to publications.

### DLR User Metrics

Katrin (DLR) gave a presentation of user metrics at DLR. She described the user metrics reported in national Copernicus platform CODE-DE, as well as the Data Download Report that is used to view metrics. The metrics of interested include the fields of activity of the users and the user categories: country of origin, user category, field of application. Katrin described the sources of user information and the reporting system architecture.

For personal data protection, rendering users anonymous for CODE-DE reporting purposes is essential; hashing user names to protect identity is one method they use.

DLR also tracks data use via persistent identifiers.

Katrin’ conclusions and discussion points are:

* User metrics tracked and reported based on project requirements
* Evaluating analysis of information held in overall User Management System and available from various ordering and download services (ETL messages, log files)
* To do – explore analysis of ‘user satisfaction’
* Review and compile possible sources of information (helpdesk, orderdesk, …)
* Unstructured information (emails, phone calls, …)
* Surveys are a possibility – not systematically done yet
* CEOS initiative welcome
* Harmonized set of metrics to track
* Harmonized user survey questions
* Technical approach for reporting to CEOS

Mirko commented that the information provided by Katrin will be very helpful when developing the best practice, especially around security and privacy; the team would welcome DLR’s contribution to this task.

### NASA User Metrics

Chris Lynnes (NASA) gave a presentation on user metrics at NASA. NASA uses the American Customer Satisfaction Index (ACSI) obtained from the annual survey of registered users of EOSDIS. He displayed examples of scores obtained by user type, and noted that free text comments in the survey are very helpful and actionable. Surveys are also given at the start or end of a webinar.

Chris described the EOSDIS automated metrics system (EMS) and provided a few results. He mentioned the Association Rule Learning which finds associations between variables; this is used by Amazon and others for recommendations, using the number of times that two items are ordered together (placed in the same “market basket”).

Of key interest to NASA are impact metrics: papers written using dataset X and applications using dataset X.

## User management

### ESA Single Sign-On (SSO) and Federated Identity Management

Marco Leonardi gave a presentation on ESA Single Sign-On (SSO) and Federated Identity Management. He described the current operational status, which is based on the SAML standard and uses a self-registration mechanism. He noted that SAML federations are currently not supported by EO-SSO. An evolution of the ESA user and identity management infrastructure is currently in place aiming at standardising the overall architecture and at integrating new functionalities.

Current pathfinder activities include a Cloud services access pilot (Q4 2017), and the ESA Earth Observation SAML federation pilot (Q2 2018).

Future plans are to move the results of the pathfinders into operations. The ESA Earth Observation Single Sign-On infrastructure (EO-SSO) will facilitate user access to satellite data and to the Exploitation Platforms’ services by supporting standard digital identities federations and will support the integration with existing research (inter)federations like eduGAIN by enabling the research communities to access satellite data and services in a standard and simplified way. The ESA EO-SSO will be able to make the federated user identification an enabler for the Exploitation Platforms in the context of the Network of EO Resources.

Esther asked if they are planning to use this for shared work spaces. Marco replied that the evolution is to cover not only SSO but other mechanisms, such as using certificates. ESA will follow standards to improve security and will make their plans available as part of the governance and policy decisions as a federation with other entities is built.

SSO is attractive, but the users’ networks need to be verified to mitigate vulnerabilities; infrastructure and security procedures cannot be reduced. Users will have to prove who they are and why they are asking for the data. At the moment they are not implementing the strictest control because of the nature of the data.

Mirko invited Marco to participate in the draft a federated management BP for CEOS.

### USGS User Management

Kristi Kline gave a presentation of USGS user management. She described the current process for user registration, which includes storing of an encrypted user profile that ensures privacy and compliance with regulations. It also includes reporting on demographics, such as how the data is used, quantity of data download, and aggregate statistics on the user’s organization. She described the architecture and showed examples of the screens.

Future plans include consideration of Open Authentication, which tracks downloads, but also needs to track utilization of data from cloud.

Mirko noted that it would be helpful to hear their results on metrics on cloud. He invited her to contribute/review document on BPs.

### NextGEOSS User Management

Juan Doval gave a presentation on NextGEOSS User Management. He began with some perspective on user requirements, and the main functionality of GEOSS user management. He described state of the art protocols for authentication and authorization, and the key performance indicators that they monitor.

Juan described the short term plan for application and services integration with an SSO federation of ESA, NASA, and OGC Testbed14. Their user management serves the integration of NextGEOSS services in community portals with SSO authentication and authorization capable to restrict access to resources. Authorization is meant for secured NextGEOSS Services and is based on scope set with default values. The administrator will be able to modify the claim values after request and approval. Scope authorization uses the user-info endpoint on NextGEOSS UM side.

Mirko invited Juan to be available to help with the WGISS BP development, especially since they are accessing IDN, CWIC and FedEO.

# Joint Session with WGCV

## Welcome and Introductions, adoption of Joint Plenary Agenda

Mirko Albani (WGISS Chair. ESA) and Kurtis Thome (WGCV Chair, NASA) welcomed the participants. They both commented on the excellence of the meetings, facilities, organization, location, and organizers. They stated that this meeting will have many opportunities for good discussion and interaction, and expectations are high given the topics for discussions. Mirko reviewed the agenda, and asked the participants to introduce themselves.

## Host Welcome Opening Address

Dr. Ricardo Galvão, INPE General Director, welcomed the participants to the meeting, and expressed regret that he could not attend all the sessions. He invited everyone to visit the INPE administration and directorate. He acknowledged that this meeting is very technical wished that the working groups achieve all their hopes.

Incoming GEO director, Gilberto Camara (INPE) also welcomed the participants, and shared some important points. Gilberto gave the opinion that since he began his work in 1978 there has not been a time more exciting than now. This results in a responsibility to ensure that organisations deliver to people to the maximum.

Gilberto stated that the major changes of the past five years are here to stay. The very existence of CEOS is magnificent, representing space agencies with different agendas, but getting together to solve tough technical problems; working together is a thing of marvel, and because of this GEO has a chance to succeed.

If you look at the questions CEOS is asking, they are getting tougher and this is grounds for optimizing. GEO is committed to give to CEOS just as CEOS is giving so much to GEO.

At this point, some of the promising activities are far from ready for any kind of standardisation. People are exploring completely novel ways to understand EO data. The way to deal with time and change of data is by thinking about it after the fact. Organizations are in a state of flux and this is a positive thing.

Gilberto expressed the hope to bring to GEO this sense of excitement – just think on-demand ARD! The achievements of the past are something the agencies should be very proud of, and he looks forward to exploring together.

## CEOS Executive Officer Report

Steven Hosford (CEO, ESA/CNES) gave the CEOS Executive Officer (CEO) report. He began with greetings from the current CEOS Chair, Philippe Brunet, and from EC. Steven gave a brief introduction of CEOS, describing functions, objectives, governing documents, and organizational structure. Steven also described the annual cycle of CEOS for meetings, meetings with GEO, and gave an update of the CEOS Work Plan. He also outlined CEOS’ institutional relationships, especially with GEO and the UN.

Steven listed the CEOS Chair priorities for 2018. He added that each SIT Chair also brings a different point of view and focus. NOAA has solicited feedback from all WGs and VCs in February through questionnaire and conference calls. Much food for thought was generated, and the SIT Chair team will digest and reformulate the data. He described the upcoming SIT-33, where WGISS will have active participation. He concluded saying that the hope of the Chair is that the working groups will contribute to making the planet’s EO programmes greater than the sum of their individual parts.

Mirko highlighted the goal of SIT for sustainability and interaction of the CEOS working groups. Determining the way forward, and asked Steven what type of interaction he envisions. Steven replied that the working groups are there to achieve the functions of CEOS; those aspects are not static, and are evolving with the changing environment of EO. One concern is how to make it clear to the agencies that the global integration is where most value can be had for the users. For some agencies this has been realized, but for others it is overwhelming. He envisions continued participation by motivated people from an organizational perspective, demonstrating that the work is contributing to something that is bigger. Mirko suggested addressing this at the SIT meeting.

Discussion continued about functioning with the virtual constellations, reinvigorating the interaction and functioning together. The VC is the final customer of the WGs, and this is a very ambitious objective. The VCs should find capacities and skills in the WGs, though there are big difference in maturity among the VCs. Kurt added that the WGs have to get better at seeing that they are providers and take that responsibility.

## CEOS 2018-2020 Work Plan

Steven Hosford discussed the CEOS Work Plan (WP), whose focus is on objectives and deliverables, associated with timescales and assigned responsibilities. The work plan is updated annually but based on a three year period and is endorsed virtually by member agencies. The themes are climate, carbon, agriculture, disasters, SDGs, water, oceans and coasts, polar, and biodiversity. The cross-cutting themes are data access, data quality, VCs, and outreach.

Steven discussed the difficulties and challenges of the process of updating the WP, adding that the iterative process is working well. There is a concerted effort to align the CEOS Work Plan with CEOS entities’ management process. Some ideas for improving the process are CEOS entity meeting scheduling and WP discussions, and improving contact with CEOS entity leads.

The SIT telecons are very useful, especially in terms of feedback; a similar teleconference with the CEO may help.

## GEO Secretariat Report

Paola de Salvo gave the GEO Secretariat Report, giving a brief history of the organization, concluding with the evolution in 2017 of the GCI to the GEOSS Platform. She listed the variety of data providers which are regional and national, and also from the private sector. GEO offers a number of components in support of the community, and from the GEO DAB can enable machine to machine connections. GEO is conscious that the number of resources available can be overwhelming to the user.

Paola listed a number of elements available to the user:

* GEOSS API
* GEOSS View allows to subset what is of interest to a domain.
* GEOSS Platform
* GEOSS Mirror for communities that need a user interface. Enables any community to replicate and have their own customizable view, but with same look and feel of GEOSS.
* GEOSS Widget
* GEOSS Yellow pages, registration process to simplify the path for data providers to come in.
* GEOSS Status Checker which provides feedback on status, alerting the data provider and user
* GEOSS-like to enable the user the provider quick feedback that was discovered from the platform.

Paola listed the standards supported by GEO, and gave a few metrics (170 brokered catalogues, 5000 different data providers, 414 million resources). She highlighted GEOSS Portal enhanced functionalities: map-based, simple and extended search, smart filters, information and visualization services, personalization/profiling, and geo-localization services. The portal is in collaboration with base map selection, and public-private partnership.

GEO follows a user-centric approach, considering various user communities, with a focus on reusability, collaboration, and user feedback. They are extremely keen to provide interoperability with other APIs, like the CEOS Data Cubes.

CEOS Brokered Catalogues into the GEOSS Platform are in the GEOSS Yellow Pages.

Paola acknowledged CEOS Participation to the 3rd GEO Data Providers Workshop in May; they are extremely excited by the strong participation. They are also eager to document usage of CEOS CWIC and FEDEO data catalogues via the GEOSS Platform.

Mirko thanked Paola for the overview of the platform, with all its interactions, adding that WGISS is working on many of the things displayed on the interfaces diagrams. It would be worthwhile to organize an interoperability session at WGISS-46 to move ahead on the interoperability. Paola said indeed, she would like to proceed. She suggested discussing this at the Data Providers Workshop where Copernicus services the flagships, and regional groups will be represented.

**Action WGISS-45-12:** Mirko Albani to work with Paola De Salvo to organize a GEO-CEOS interoperability session at WGISS-46 in order to identify how WGISS can cooperate with the interoperability goals of GEOSS by end of May, 2018.

## CEOS Systems Engineering Office Report

Brian Killough gave a presentation on the CEOS Systems Engineering Office Report. Brian began with a COVE Tool update, and showed the 2017 coverage (all instruments) of CBERS-4 over Brazil using the COVE tool. The user can click on any pixel to get more information.

Brian continued with a map of the ‘CEOS DC Road to 20’ plan: four Data Cubes are operational, 11 are under development, and 28 countries are under review or have expressed interest. He reported that there has been a lot of activity and various levels of implementation, and the team is moving rapidly forward. Data Cube plans include new deployments, collaborations with Google and Amazon, the Swiss Data Cube Hackathon, IGARSS Conference, new technical additions of Jupyter Notebooks and web-based UI tools, and new user applications and algorithms.

Current WGISS-SEO collaborations are:

1. Maintain and expand the connections from satellite mission archives to the COVE tool.
2. Investigate approaches for on-demand DC creation using cloud-based mirror sites, or other data sites for discovery, processing, and ingesting of DCs to support global users.

Rob commented on a discussion of “on-demand”, where a joint technology testbed was proposed. The Python client can be considered at that time. Targeted datasets are Landsat, and Sentinel.

One of the areas where there has been no progress is with Sentinel-2 data; the methods of access are inconsistent.

## WGISS Overview

Mirko Albani gave an overview of the activities and expertise of WGISS, focusing on the objective to make data discovery accessible and usable to users. He highlighted a few success stories and cooperation with other CEOS and GEO groups.

## WGCV Overview

Kurtis Thome gave an overview of the activities and expertise of WGCV. He explained that the working group is organized into six subgroups, covering the general topics of validation metrics, protocols, traceability, and pre-launch and post-launch characterization. Some of their activities include:

* The CEOS WGCV Land Product Validation (LPV) has established a framework with the aim of independent validation and consistent uncertainty reporting across products as main output.
* The Atmospheric Correction task group is working toward better understanding of the different uncertainty contributors.
* A network of instrumented sites is dedicated to the radiometric calibration of EO optical sensors developed in the IVOS Subgroup. There are four automated test sites around the world. The plan is to accept these for the RadCalNet sites and then open them to the public. Sixteen groups volunteered to participate.
* SI-traceability, developing good practice approaches to ensure laboratory and field measurements are interoperable.
* Solar Irradiance spectrum choice can lead to differences in comparisons between sensors. WGCV-accepted solar irradiance spectrum will be distributed on the CEOS Cal/Val Portal.
* Level 1 interoperability goal is to develop an initial recommendation of a community reference in collaboration with GSICS. Must be good and highly accurate before moving to Level 2.
* Fiducial reference measurements to determine how to ingest in a format that is useable. Data can be uncertain, but the user needs to know the level of uncertainty.

Summary:

* WGCV has seen an increase in cross-cutting activities amongst its subgroups
* Similarly, CEOS is seeing an increase in cross-cutting activities amongst its Working Groups and Virtual Constellations, and they are learning to address those.
* Limited resources point to the benefits of cooperative efforts amongst the various WGs and VCs
* Joint meetings allow for discussion of common areas and topics in the ongoing/planned WGs activities that would benefit from cooperation between the two groups: Describe and share available facilities, expertise, resources, technical information and documentation, and topics of common interest and identify synergies focused on common tasks.
* Next set of talks will hopefully lead to clear paths forward to produce measurable benefits to CEOS.

Makoto asked how to access their reference data. WGCV replied that they are improving their access, and hope to learn how to do a better job at this meeting. There is no unified method, but rather individual datasets. All of this information can be found on the CalVal portal or the WGCV website.

Michael noted that the IDN does have some registrations of entries done two years ago, for the LandNet sites and the RadCalNet sites; but they need to be updated.

Mirko asked where their good practice documents are; they can be found on the LDPV portal.

**Action WGISS-45-13:** Michael Morahan to follow up with Greg Stensaas to update the test site information in the IDN by end of June.

## Overview of WGISS/WGCV Joint Effort

Mirko Albani and Kurtis Thome gave an overview of the WGISS/WGCV joint effort. The scope of the joint session is to discuss common areas and topics in the ongoing/planned WG activities which could benefit from cooperation between the two groups. There will also be discussion of common areas and topics in the ongoing/planned WGs activities which could benefit from cooperation between the two groups.

Identified topics to be discussed

1. Data Formats and Interoperability in the framework of FDA
2. Quality Indicators in Discovery Metadata
3. CEOS Data Cubes and CEOS Test Sites Data Access in support of WGCV Activities
4. Standardization and Best Practices (e.g. ISO 19159-3)

Philippe Goryl (ESA) suggested that all these topics be underlined by ARD, as many struggle with the definition. Medhavy pointed out that the LSI VC has put together specifications for three categories of products: the CEOS ARD for Land for the optical. For radar it is more complicated, there may be more than one. The SAR subgroup is looking at the backscatter definition. He also noted that for CARD4L WGCV was consulted.

Albrecht mentioned that for ocean and atmospheric products the VCs are being polled for what they want for ARD; the LSI VC is the most mature so leading the way.

## Data Formats and Interoperability in the framework of FDA

Medhavy Thankappan (GA) gave a presentation on data formats and interoperability in the framework of FDA. Medhavy introduced the project for Ensuring ARD Interoperability, which has the following outcomes:

1. ARD correction parameter sensitivity: Determine the key set of common parameters between correction approaches to minimise artefacts in reflectance and downstream processes
2. Temporal stability: comparison of measurements from corrections through time (GA NBAR method vs LEDAPS/LaSRC )
3. Field validation: comparison with in situ measurements (historical and current)

Medhavy listed the ARD correction parameters and described the test sites. He noted that the parameters that affect the correction are solar angle, aerosol and water vapour content and the bidirectional reflectance distribution function (BRDF). Sensitivity analysis includes solar angle, aerosol change, water vapour, and BRDF, and temporal stability.

Medhavy described a comparison tool between the GA product and the USGS product. There is a seasonal component to the differences. In the near infrared. Red, and similar trends between coastal and agricultural regions. He also listed the key elements of validation.

Robert Woodcock noted that the underlying formats are less of an issue, except when using a specific GIS format. Conversion processes handle these.

The CEOS objective is to ensure a seamless access and usability of data from multiple sensors including tools to fuse/use data from multiple sensors:

* Data/Sensor formats interoperability: e.g. CEOS Format update
* Joint contribution on ARD definition
* Determining uncertainties for Analysis Ready Data (and methods for communicating those uncertainties)

Nigel Fox commented that for the user, Landsat-8 needs to be the same regardless of the provider; in addition, knowing the uncertainty is fundamental.

Medhavy noted that they are trying to characterize the differences between sensors. Fundamental between two sensors is their spectral response function.

Medhavy proposed that WGCV and WGISS address the following:

* Different data formats such as CEOS, netCDFx, HDFx, ENVISAT, GEOTiff, MPH/SPH/DSR, HRPT🡪 alignment of data format allows data interoperability in terms of data comparison (visualization, geolocation, etc.).
* Different data content, such as ARD 🡪 alignment of content allows data interoperability in terms of data comparison and data management for time series etc.
* Full Interoperability Process:
  + ARD specification available for all sensors/areas of interest with a joint contribution for ARD definition, determining uncertainties for Analysis Ready Data (and methods for communicating those uncertainties)
  + Cost/Benefit Analysis that considers the product owner’s cost and the final user needs, and consider a simple converter or processor improvement.

Albrecht von Bargen (DLR) agreed that a sharp definition of ARD is needed, and this might be a contribution from both working groups.

Medhavy noted that both fusing and comparing data must be considered. Esther commented that for DCs the huge volume of data affects the requirements, and it would be an interesting case study.

Robert said that the NetCDF header cannot handle a certain level of complexity. There is no need for debate about file formats; this problem is solved by a file converter. But a debate on the content of the measurement corrections and how to communicate the uncertainty is what is needed.

For a CalVal test site – per pixel can be had if needed. What is the best methodology for taking the data quality and fusing it with the data stream?

Greg Stensaas commented that it all comes down to a way to document the steps of interoperability from beginning to end, in a broader sense to the user community.

The fundamental question is how best to incorporate the uncertainty and provenance and how to educate the user community. What is completely clear is CARD4L. Maybe that is a good starting point.

## Quality Indicators in Discovery Metadata

Nigel Fox (UKSA) gave a presentation on Quality Indicators in Discovery Metadata. The objective is to ensure quality and uncertainty information availability (discovery and access) for users. This requires

* Definition and implementation of Quality Indicator at collection level (quality, fitness of purpose) or at the granule level (for end users). Higher level product should have an indicator for the user, derived from the instrument and the processes it has gone through, and an estimate of uncertainty being introduced.
* Definition of Quality fields in discovery Metadata for the key sensors and products following the QA4EO guidelines. Include validation metrics (possibly calibration metrics and sensor performance indicator). Make the reference to validation guidelines into the metadata information. (WGISS-41)
* Define cross-validation projects across ARD products, reach out to commercial providers to encourage their participation in ARD, and promote and enable discoverability of ARD datasets. (WGISS-44)

Within QA4EO it was emphasized that fitness for purpose does not have to be the best; it just needs to be assessable by the user.

Nigel reported that a survey was conducted to discover which quality information the data users want; there were over 100 respondents.

Nigel described quality indicators and gave detail information per parameter for climate. He also described the delivered product of RadCalNet, and a portal with discovery QA metadata.

What might need to be done? QI information needs to be tagged at all stages of product (collection, granule, and higher level products) with a linked chain backwards. It should also be determined what QI available, how comprehensive and what level of validation is there. Establishing coefficients to harmonise L1 and scene dependent L2, and determining the minimum for type of applications and ease of assessing is useful, as is robustly linking evidence/documents to data sets.

Comments:

* Is there a QI that is usable in the discovery metadata?
* The difficulty is identifying how many parameters you want to define.
* The guidelines are well defined, but not the details.
* It will always be specific to the application (thematic).
* How to include the quality indicators is known: the QI field, and the related URL to do it at the collection level. Some collections have many variables included.
* How to include quality is known; what is not clear is how to describe it to make it usable to the end user.
* Quality is defined by the end user. Those parameters can be supplied by the provider, but it is up to the end user to define the quality.
* If the parameters are provided, then empower the users to define their own quality.
* The methodology needs to be expandable to some unknown future requirement. Having a straightforward test case would be very helpful.
* QI is the uncertainty of the parameters. Some aggregation can be provided in a test case.
* The objective is to define a metadata that provides the set of parameters with uncertainties, so the customer can make a judgement on the usability.
* It should be at the per-pixel level.
* The challenge is to put the expressive power to an environment where you can express all the parameters.
* Does it make sense to start with one of the CEOS DCs, or with CARD4L? The SST is a mature option. SST is well solved, could work, but applying it to another domain is another matter. SST is attractive because it is a global community.

## CEOS Data Cubes and CEOS Test Sites Data Access in support of WGCV Activities

Greg Stensaas gave a presentation on CEOS Data Cubes and CEOS test sites data access in support of WGCV activities.

CARD4L will require a lot of coordination between the two WGs. Full exploitation of EO data requires critical user feedback informing current implementations and future directions.

Recommendations:

* Focusing on WGCV Land subgroups (IVOS and LPV) and related key GEO applications. Potential for SAR subgroup scenario later.
* Obtain data over key test site areas (WGCV). Push process via CEOS SIT/ LSI VC/ WGCV/WGISS. Define and list primary sites and secondary sites (super sites, pseudo invariant sites, LandNet/RadCalNet sites, LPV defined key sites).
* Make data easy accessible (WGISS).
* Utilize Moderate Resolution Interoperability as test platform in the data cube (WGISS/LSI VC – development).
* Test example use case(s) with WGCV as user. Define key process events.
* Test example use case(s) with key application users. Define key application user test scenario (GLAM).
* Compare data via visualization and trending/monitoring tools (GEOCube and WGISS - development) (User - WGCV IVOS/LPV).
* Data Cube definition and plan (WGISS).
* DC test site analytics environments with ready-to-run Jupyter notebooks available on, say, AWS so people can spin them up on demand. (WGISS)
* Detailed documentation related to interoperability WGISS/WGCV/others.
* “CEOS Data Cubes and CEOS Test Sites Data Access in support to WGCV activities “
* WGCV/WGISS goal:
  + specific test case(s) of WGCV-type data
  + Using Cal/Val Test sites
  + Use as example(s) to show how future data cube efforts could be standardized. Involve standardization of LandNet, MRI, and ARD interactions.

Robert commented that there are two key topics: access and analysis. Would it be useful to have an on-demand analytics environment; this is what FDA is intended to provide.

The LPCS takes the test site information and, based on the size, re-projects.

Cindy Ong suggested that it would be nice if ancillary data was also accessible in these test sites. The LPV supersites are being used continuously and many of these already have it, and the user community is already there.

Fernando Camacho (CEOS LPV) said these sites are very adequate and have a permanent perspective, so it would be very appropriate to start with them. At ESA they are trying to identify the key supersites in Europe.

## Standardization and Best Practices

Cindy Ong gave a presentation on standardization and Best Practices (e.g. ISO 19159-3). The objective is contribution for the CEOS Best Practice/White Paper. She listed the following activities:

* Support and collaboration on ISO 19159-3 “Geographic information -- Calibration and validation of remote sensing imagery sensors and data – Part 3: SAR and InSAR.” (WGISS-43)
* Develop a White Paper on Recommendations for improving Data Quality in Metadata (WGISS-41)
* Support evolution of Data Management principles (Quality principles)
* Support WGISS to provide guidelines for the metadata for ARD to ensure the information needed to describe a measurement is available. (WGISS-41)
* Develop Best Practices on data reprocessing – including current and future practices (i.e. on-demand ARD Processing) (WGISS-41)
* Support Metadata Requirements for Quality (Quality fields): Review and comment on GEOSS Data Quality Requirements (DDQ -Documentation on Data Quality). Include validation metrics (possibly calibration metrics and sensor performance indicator) Make the reference to validation guidelines into the metadata information. Develop a White Paper on Recommendations for improving Data Quality in Metadata.
* Support evolution of CEOS Best Practices and relevant standards:
  + Review of Data Management and Stewardship Maturity Matrix for the Quality area. The result will be provided as input for GEOSS-DMP Implementation Guidelines improvement (Quality principles: DMP-6).
  + Support and collaborate on ISO 19159-3 “Geographic information -- Calibration and validation of remote sensing imagery sensors and data – Part 3: SAR and InSAR”.
* Support WGISS to provide guidelines for:
  + ARD metadata definition for identified areas of interest and/or key sensors.
  + Data Reprocessing Best Practices development – including current and future practices (i.e. on-demand ARD Processing).
  + Data Quality in Metadata white paper and recommendations definition.

Kurt asked if the SAR example can be included in the test; it has already been approved.

Greg said that for one of the standards (ISO 191959-1) WGCV provided comments, but there are still updates that need to be done. Contribution will result in a standard that the agencies can be comfortable with.

Nigel warned that standards are being set before the sector is mature; need to wait for maturity before adopting standards.

The WGs should be defining good practices so they can become the basis for future standards.

## Summary on Joint Interaction and Additional Opportunities; Joint Recommendations to CEOS and GEO

Mirko Albani and Kurtis Thome presented a summary of the session, making joint recommendations to CEOS and GEO.

Discussions covered common areas and topics in the ongoing/planned WGs activities which could benefit from cooperation between the two groups:

* Ways to describe and share available facilities, expertise, resources, technical information and documentation
* Topics of common interest with natural synergies
* Defined tasks to define a way forward
* Agreed to monitor progress through periodic joint calls

The following objectives provide a workable summary:

**Objective: Ensure users a seamless access and usability of data from multiple sensors including tools to use data from multiple sensors**

* Example case for surface reflectance showing the range of possible challenges in creating a consistent time series across multiple processing schemes.
* Data format is not the biggest issue to solve related to interoperability.
* Format conversion is pretty well understood and WGISS already provides that service.
* Recognize that the definition of ARD is necessarily broad; this should not prevent progress on understanding how to communicate uncertainties and data quality.

Task - ARD uncertainty: Work to define a set of quality/uncertainty parameters to include within ARD and determine an initial method to incorporate these along with their provenance into the data stream, including education of the data community on how to provide provenance information.

**Objective: Quality and uncertainty information availability (discovery and access) for users**

* Measurement equation must be embedded in a way
* Per pixel too hard to handle in a continuous operational mode.
* There exist a set of key quality metrics that are useful and indicators.
* QI should be tagged at all stages with a link to the earlier stages and is locked into the future.
* Discussion on where to go forward – QI is well defined but concrete cases will always have different levels and flags.
* Describing QI is not easy and more difficult to do it for the user readability.
* Keep in mind that quality is defined by end user – fitness for purpose is defined by end user.

Task - QI Test case:Evaluate whether SST would provide a suitable test case for QI development that can act as an expressive case for QI access and implement this or an alternate case as determined through this activity.

**Objective: Data Cubes (or similar stacking of data over time for specific locations) test cases**

* A range of datasets already exist and are available through the Cal/Val portal such as LandNet; plus QA4EO activities; and LPCS.
* Need a way to get those data more available to users.
* Recommendation – Focus on IVOS and LPV and key GEO applications including obtaining data over key test sites (WGCV) and pushing this process to CEOS SIT/LSI VC/WGCV/WGISS.
* Simplify the problem from a WGISS perspective and concentrate on analysable and get an on-demand analytics environment to do what users want.

Task - LPV supersite data access test case:LPV supersites are a good test case related to validation with a broad user community. Objective is to provide an improved user access to the LPV supersite data sets coupled with analytics tools to provide improved results for the users.

**Objective: Contribution on CEOS Best Practice/White Paper**

* Discussion from 2016 is still valid today.
* Work on actions from 2016 still remains.

Task - Revisit and update WGISS-41/WGCV-40 actions:Define a point of contact within WGCV to work with WGISS to reinvigorate the actions of WGISS-41 in terms of more recent CEOS activities and move towards closure of these items to provide a good practice white paper for inclusion on the CEOS website.

Discussion:

Providing per-pixel to those who need it is manageable. From per-pixel to collection the uncertainties increase. Per-pixel is too hard to handle in a continuous operational mode. The per-pixel case can be done with the SST example.

Kurt commented that he found this session useful, with tasks that will forge a better collaboration.

Mirko suggested that the coordinators of these four sessions also coordinate the tasks.

Kurt noted that Philippe’s concern about focusing on the DCs in the test sites task will be addressed.

**Action WGISS-45-14:** WGISS Exec to organize bi-monthly telecon with WGCV to formalize as actions the tasks identified during the Joint Session, and to review status of these. First telecon in June.

# Workshop on Future Data Architectures (FDA) and Introduction of Copernicus Data and Information Access System (DIAS)

Mirko Albani introduced the Workshop on Future Data Architectures (FDA) and Introduction of Copernicus Data and Information Access System (DIAS).

## FDA Context

Steven Hosford (ESA/CNES) gave a presentation on the FDA context. He began noting that a critical objective of many space agencies is to position themselves to reinforce this new complementary activity and ensure EO becomes part of everyday life is. He described the background of FDA within CEOS, and gave its status at the CEOS Plenary 2017.

Steven discussed Analysis Ready Data (ARD), CEOS Analysis Ready Data for Land (CARD4L) and interoperable products within the evolving EO ecosystem. He also discussed Medium Resolution Sensor Interoperability.

The functional blocks of FDA can be characterized in a four layer model with a data resources layer, a data generation layer, an exploitation layer, and a platform services layer. If the layers are characterized properly the components can be handled.

Chris Lynnes expressed appreciation for the users vs. compute plot placing FDA ecosystem elements in context, and suggested that it would also be useful to map or graph the connections among the elements. Steven indicated that the plot provided as an example was indeed only one example of the way data gathered in the inventory process could be presented, and there would be many other sets of parameters that could be displayed to provide insights.

Robert commended the effort to mature the language involved in the FDA conversation. Two additional comments on the 4 layer model are that need to make the CEOS shared components, the bits that make FDA a CEOS common community effort (e.g. IDN). The interfaces and connections between components need to be identified to show how things need to cooperate. Steven agreed that if he was referring to specific CEOS common efforts (such as WGISS provided discovery functions), those could be identified as a layer, however unspecified common CEOS community efforts are more difficult to represent as they may be multi-form and therefore potentially in several layers.

## Exploitation Platforms and Common Reference Architecture

Cristiano Lopes (ESA) gave a presentation on exploitation platforms and a common reference architecture. He described the terminology, and explained ‘Space 4.0i’, which describes the way ESA will play its role as a space agency for Europe. Space 4.0 combines the described global situation of space developments with the ‘i’ standing for an ESA-specific interpretation of the tasks.

Cristiano described DIAS, which is a resource tier layer implementation for European access to Copernicus data and information. He described the infrastructure and architecture.

Cristiano concluded with the implementation status.

## ESA Perspective on Data Cubes

Andrea Della Vecchia (ESA) gave a presentation on ESA Data Cubes. He highlighted the number of missions and the volume of data that needs to be organized for provision to the end user. ESA recognizes that a combined analysis of these data streams creates a huge potential to better understand the trajectories of land and marine ecosystems, atmospheric components, and the interactions of these spheres. Andrea posed the following questions:

* Can data cubes support the scientific community to adopt a more holistic approach in understanding land-marine-atmosphere interactions and the role of humans?
* Can they provide an innovative data access mechanism to ease the life for users from outside the space community?
* Can data cubes support the drilling through long time series of different satellite series more efficiently?
* Can data cubes support the merging of different information sources (space, airborne, in-situ, GIS) more efficiently?

ESA has experimented data cubes with two different levels of data:

“Signal measurements” (e.g. spectral radiance, backscatter  
“parameters” (e.g. land surface temperature, soil moisture, biomass,)

Cristiano discussed dynamic generation of parameter layers and dynamic gridding, as well as use cases of the mission owner and the commercial data/information owner.

Cristiano showed examples of data analysis by clicking on a pixel and seeing various time series and specific comparisons. He gave as example an investigation of drought in eastern Africa in last 17 years, with time series of vegetation index, precipitation, and soil moisture. The full collection is more than 5 TB, but the user can use less than 1 MB to identify anomalies and then drill down to the detail data for more study.

Cristiano also showed the federated exploitation infrastructure for allowing data and services discovery, EO products visualization, data processing, and data access/exploitation/storage.

Homero asked if Jupyter notebook is similar to Google Earth engine. Cristiano replied that the approach is quite similar from a functional point of view.

Simon Reid commented that this method is insecure because of its power; it gives access to the root and it is not easy to protect against that. However, this is just a pilot for proof of concept.

Richard wondered if this development will be open source; Andrea replied that it will be released but they are in discussion on how to make it completely open source.

## Sentinel-2 ARD: Making the Process Operational

Federica Moscato (UKSA) gave a presentation on Sentinel-2 ARD: Making the process operational. She began with a description of the Satellite Application Catapult, which is part of the collaboration with Sentinel. She also discussed the Sentinal-2 data access service (Sedas), the Sentinel-2 ARD project, and the Atmospheric Radiometric Calibration of Satellite Imagery (ARCSI) functionality underlining the cloud masking, which is free open source and multi-sensor.

Frederica described the Wales-UK DC, which is an instance of the Open Data Cube hosted by UKSA.

Future work involves creating CEMS tools, and the IPP Common Sensing Project.

Mirko commented that in the past there was also work to ingest SAR data to the cube, and wondered if that is continuing. Frederica replied that the current work is in the optical, but SAR is in the future plan.

## Open Data Cube in Uganda

Simon Reid (UKSA) gave a presentation on Open Data Cube in Uganda, a project of the Drought and Flood Mitigation Service. He gave the context of the project, its status, progress made, and focus in 2018. He described the system architecture, and process flows and resource management.

Simon described multiple data cubes for scenarios where they would like to connect different DCs together; this is not possible at the moment, but should be in the future. These scenarios would allow for reduced duplication and when adjacent different countries are collaborating. Another use is satellite/instrument providers providing ARD, regional caching of ARD, and where remote/bad/intermittent network connections.

Simon described the development and interest areas of the Open Data Cube, and summarized the Uganda DC, which is a complex scenario with many actors, processes, data types that helps capture and manage that complexity. Open Data Cube and ARD is a first step. The derived data are important and real power arises from combining diverse datasets. ODC is integrated into a wider cloud-agnostic platform, managing the processing chain, compute resources, security and billing, etc. ODC is also useful for enriching flow of information between stakeholders and building an accessible store of key environment information as a national asset

Mirko asked if the platform is operational and would it be possible to get a link. Simon replied that they are still in the development phase, with an instance that is currently being evaluated. A beta version is planned for October, and an operational version should be ready in about a year.

## Horizontal Scalability for the Colombian Data Cube

Pilar Lozano-Rivera and Harold Castro gave a presentation on horizontal scalability for the Colombian Data Cube. Pilar discussed the timeline, advances and results.

Harold continued with a description of the current implementation, and the architecture of the web portal, including its components. The new distributed architecture was compared to the old architecture and has many new advantages; it is being tested for scalability, ingestion performance. They have around 10-20 users; in the next semester they will open the DC to environmental institutions in Colombia, and at the end of the year should have 100-150 users.

Harold listed the following conclusions:

* This architecture may be used as reference architecture for other institutions and research groups deploying customized version of the ODC.
* The distributed architecture defined allows to scale the data cube horizontally using a cluster of Cube Servers/Workers, avoiding the vertical scaling restrictions of single server cube installations.
* The installation process of the distributed architecture required a considerable effort
* Performance tests with a larger number of Cube Servers/Workers are required to identify possible architecture bottlenecks as more Cube Servers/Workers are added.
* Tests with other types of algorithms are required to evaluate the behaviour of the architecture with other algorithms.
* Tests to evaluate the scalability of other components of the systems may be of interest, in particular the ability to scale the bulk ingestion process.

Mirko asked if they plan to use data from other missions besides Landsat. Harold replied that they want to ingest Sentinel 1 and 2 in the short term.

In response to a question from Mirko, Harold stated that they began with virtual machines on Amazon.

Steven commented that in mid-2018 there will be a surface reflectance product released by ESA for Sentinel-2.

## Copernicus DIAS

Martin Ditter (EC) gave a presentation on Copernicus DIAS. He described the objectives, core principles and structure of DIAS, and described the Copernicus challenge. Martin listed DIAS mandatory functions in terms of data holdings are Sentinel data, Copernicus Services Information, and a selection of third party data. Services are discovery and view services, basic downloads, basic programming / application environment, open source tools, and basic user support services.

Other DIAS functions for commercial offers are hosting of Third Party data, an advanced development and application environment, ICT resources, advanced/commercial tools, third party business hosting, and a marketplace for Third Party value- adding services.

Martin described the DIAS functional context and the rollout schedule with operational launch in June, noting that services and offers evolve as customers demand. He listed DIAS providers, and how DIAS fits within the EO ecosystem.

Use cases are designed to maximize return on investment for institutions and industry. Several use cases are also in discussion for progressive integration, support to on-boarding and to development, creation of a marketplace, and integration of industry data and services.

Richard commented that interoperability in a wider European context is very important, so connecting with other entities should be pursued. A scientific community might want something very specific, so it is best not to restrict systems from providing for users that are not yet known.

## GeoHazards Exploitation Platform (GEP)

Jolanda Patruno (ESA) gave a presentation on GeoHazards Exploitation Platform (GEP) and BIOMASS Missions Application Platform (MAP). She gave a description of the GEP, and of the advanced features and capabilities being developed for version 2. Jolanda listed the data that will be available (globally) and the pilot and prototype projects that are underway.

## Proba-V Mission Exploitation Platform (MEP)

Andrea Della Vecchia (ESA) gave a presentation on the Proba-V Mission Exploitation Platform. This is an ESA pathfinder project to integrate correlative data and derived products addressing the wider vegetation user community with the final aim to ease and foster the use of Proba-V data. He described the architecture and described the tools for data access and analysis.

Andrea also described the ESA-NASA multi-Mission Analysis Platform for improving global aboveground terrestrial carbon dynamics.

Richard noted the MAP is operational, and asked how they solve the issue of security. Andrea replied that they rely on user authentication and on a support team in case of attack.

Homero asked what is the difference with what was shown earlier. Andrea replied that this is another instance (this one is a federation) that connects both to access each other’s data; an exploitation platform is focused on a specific theme. This work is to have all of them talking to each other.

## USGS Experiences with Cloud Hosting

Kristi Kline (USGS) gave a presentation on USGS experiences with cloud hosting. USGS is working on modernizing processing, access, and distribution of Landsat data, and she listed the steps. She described the proposed architecture.

Kristi reported that USGS is re-imaging everything, and explained that data in the Cloud is not compressed. They are working toward Cloud Optimized GeoTIFF ready for immediate use in the cloud; not necessary to download (use the cloud systems). USGS is also working on getting added tools and IT security aspects.

Kristi described the Landsat Image Assessment System that is moving data into Redshift; initial tests show that queries are significantly faster.

Mirko commented that it would be interesting to hear more technical details at WGISS-46. Kristi said she could send out a study of data formatting in the cloud.

## CODE-DE Platform

Albrecht von Bargen (DLR) gave a presentation on the Copernicus Data and Exploitation-Deutschland (CODE-DE) Platform. Albrecht explained that CODE-DE is a system for access and distribution with portal search and access, geo-client and data download. It allows open/public processing and third party processing, and he described the dedicated third party processing.

Way forward is the successful installation of further CODE-DE geo-client components, integration of Sentinel-3, Sentinel-5P and RESA data, and integration of further tools.

Makoto asked about the users; Albrecht replied that the users are research and public institutions. There is also a cloud implementation for private platforms.

## How Cloud Computing Can Help Scientists

Chris Lynnes (NASA) gave a presentation on how cloud computing can help scientists. He remarked that cloud benefits are only achievable when the processing is in the cloud. This can be much faster, has access to large amounts of data, is cheaper, and can be used to gain deep learning. This is a new processing paradigm and there are many of packages and frameworks to help with this. Cloud computing is resilient but one cannot rely on persistence.

Chris suggested that to get started scientists can take advantage of free tiers for learning, and there is a lot of online training. Learning Python is also helpful.

## Cumulus: Cloud-based Archive for NASA’s Earth Science Data

Katie Baynes (NASA) gave a presentation on Cumulus: Cloud-based Archive for NASA’s Earth Science Data. She described the function of EOSDIS, which is to capture and clean (satellite, airborne), process, archive, transform, and distribute to research, applications, and education. EOSDIS has many data centres and large data volume.

Katie described the Cumulus architecture and the phased methodology. She described the transition from small system, then gradually expanding their reach.

Cumulus prototyping worked on distributing responsibility; to this end they created a set of migration teams and a core development team. The concept is to promote group ownership and buy-in, and the vison is that all the data centres are running the same version. For end user migration first they moved the core applications to the cloud. Data has to be in the cloud before processing in the cloud can begin.

## Best Practices for Cloud-Native Architectures

Dan Pilone (NASA) gave a presentation on Best Practices for Cloud-native architectures which are a fundamentally different way of building applications. Dan described the Global Imagery Browse Service (GIBS) in the cloud service swap. By leveraging AWS services, processing was reduced significantly; architecturally there is one fewer layer.

Managed services drive innovation; there is greater segregation of functionality and movement toward services over monoliths. Data can be generated at scale in AWS and placed in accessible buckets, avoiding massive data moves. Ingest, archival, validation, processing, etc. can scale dynamically based on incoming data streams, reprocessing needs, etc.

AWS has very low internal latency. An on-premises implementation showed consistent performance during load testing vs. more sporadic latencies in AWS. It is important to involve security from the very beginning and to layer security throughout the architecture. He added that modelling total cost of ownership is extremely complicated, especially since egress is expensive.

Dan summarized with the following:

* Enable cloud native architectures by strongly preferring cloud services
* AWS has very low internal latency
* Involve security from the very beginning
* Modelling TCO is extremely complicated
* Explore alternative architectures for possible cost savings
* Go hands-on quickly
* Incorporate operations’ needs

Regarding the questions of becoming locked in to one vendor, it was felt that the technologies are common across all the vendors, and re-architecting it for another vendor provides an opportunity to learn.

## NASA’s Earthdata Cloud Analytics Framework

Chris Lynnes (NASA) gave a presentation on NASA’s Earthdata Cloud Analytics Framework. The key questions is how to support user analysis of very large data volumes? One solution is data-proximal analysis. This enables big compute next to big data, encourages user adoption of cloud for analytics, and provides maximum analytics capability at minimum cost. It uses capabilities within NASA more effectively and efficiently and leverages analytics capabilities of external partners.

Chris listed key features, guiding principles, and the architectural concept.

Mirko asked if there a timeline for the project, and Chris replied that it is not at that stage yet.

## Vietnam Data Cube

Matt Paget (CSIRO) gave a presentation on the Vietnam Data Cube. He described the project background and timeline, with the official release in March. Matt described the user interface and displayed a few images.

Next plans include improved data management tools, full coverage with S-1, UI customisations, and application development. With the domestic and international partners they plan capacity building, strategic positioning, and application partnerships.

Makoto suggested that he speak with Shinichi Sobue (JAXA) about other satellites; there may be other useful ones. Matt added that SAR-type is preferred because of the cloud cover.

## CNES FDA Status

Richard Moreno (CNES) gave a presentation on the CNES FDA status. He described the vision of FDA, which can be handled at two levels: Interconnection of distributed systems/platforms, and at the level of the platform. Richard listed and described four FDA examples:

* Research Infrastructure “Earth system”
* PEPS : French Copernicus Collaborative Ground-Segment
* GAIA ground-segment
* CLS ground-segment

Richard noted that Hadoop does the management of huge data volumes, efficient parallel processing, efficient resource sharing, good reliability, scalability, robustness to hardware issues. However, there is poor performance monitoring, a very complex configuration, and requires a dedicated team of experts.

## INPE’s MODIS Data Cube

Lubia Vinhas (INPE) gave a presentation on INPE’s MODIS Data Cube, since big data requires new conceptual views to meet the user needs of analytical scaling, collaborative work, and replication. INPE’s motivation is to work with image time series.

The e-sensing research uses machine learning with high-dimensional input space resulting in machine learning predictors. Current applications of image time series select ‘best pixels’. They use all of the data; nothing is discarded. SVM classifiers are simple to train and robust to outliers. Function estimation by repeated approximation results in deep learning.

Lubia showed an example of deep learning training. They are having good results in terms of accuracy, and have learned that the best samples provide the best results. SITS is an R package for image time series, and is available on GitHub.

The team has learned not all data cubes are alike: Google Earth Engine is organized as 2D images over multiple machines (good for space-first, not fit for time-first). INPE’s Data Cube is currently organized as temporal bricks (good for time-first, not so good for space-first). Can a Data Cube be organized to meet both space-first and time-first efficient processing? Lubia showed an example of mixed Landsat8-MODIS bricks for Brazil.

INPE has another project to work for the five biomes of Brazil and is refining the timeline.

Robert commented on data cube optimization for time and space series. The way INPE’s operates removes some of the problems; the chunks are quite small and encouraged from different physical servers.

Chris noted that he was intrigued by machine learning and wondered about best practice for this. Lubia said this is fully documented and the results are being published.

Lubia concluded saying that they still have the data on their local small server, and moving to Amazon requires just a few modifications; the R package contains the algorithms. FAO has a platform in a front end to Amazon, and INPE can use it for some experiments.

## Summary of Workshop

Mirko concluded the workshop saying that this provides a lot of material that will help address the CEOS actions assigned to WGISS. It is hoped to work together for designing this common future architecture.

# Agency/Liaison Reports

## CONAE

Homero Lozza gave a report on the National Space Agency – Argentina (CONAE). He listed their satellite missions, noting that the current agency focus is on the SAOCOM missions. SAOCOM 1A will be launched September 2018, and 1B in 2019. The instrumentation includes SAR L-Band (1.275GHz) with several modes (strip map, Top SAR) and with interferometric capability. The satellite has a 16-day revisit pass (8 days for the constellation).

The main driver for the SAOCOM is soil moisture for agricultural and hydrological applications. A telemetric network for in-situ data will enhance the analysis. A web portal of the products has been developed.

Mirko asked if the satellite was designed and built in Argentina, and if other satellite data was also being used. Homero said that yes, the satellite was designed and built in Argentina, and that they are using SMOS and SMAP as well.

## CSIRO

Robert Woodcock gave a presentation on the Commonwealth Scientific and Industrial Research Organisation (CSIRO). He described the organisation’s objectives, investments, and capabilities, the sample target applications of NovaSAR mission that will be launched in June, and the CSIRO Earth Analytics Industry Innovation Hub.

Robert also mentioned CSIRO’s efforts on the CEOS Open Data Cube, a scalable platform for EO analytics for global impact, from EO science through a network of data cubes and CEOS engagement.

There was discussion on the link between CSIRO’s data cubes an others, the number of users, and their cloud-based training environment with play space.

## ESA

Mirko Albani gave a report on the European Space Agency (ESA). He began with description of the growing constellation of Copernicus dedicated missions, and described the Sentinel free data access for all users to Sentinel products (most recent as well as complete long term archive). He also described the ESA Data Hubs, and data access and dissemination statistics.

Mirko described the upcoming Sentinel 3 launch event, and the upcoming Living Planet Symposium.

Makoto asked if Living Planet connected with UNESCO; Mirko replied that it is not.

## Geoscience Australia

Medhavy Thankappan gave a report on the Geoscience Australia (GA). He described the Digital Earth Australia (DEA), which is GA’s infrastructure for analysis of EO. The DEA map application (currently in Beta) showcases data products and analysis for EO data.

Medhavy described the Open Data Cube, and the Second Open Data Cube Conference, and the work on the continental scale validation of surface reflectance, the national spectral data repository, the interoperability of multi-source Analysis Ready Data, and the Corner Reflectors for SAR calibration.

## INPE

Lubia Vinhas gave a report on the Brazil National Institute for Space Research (INPE). She announced the upcoming launch of CBERS 04A, and described its cameras. She also discussed the upcoming launches of the Multi Mission Platform (PMM) Earth Observation satellites for forest monitoring.

Lubia gave details on the TT and C and image reception activities, with extensive work between space and ground segments. She also discussed the comprehensive work in preparation to CBERS 04A and Amazonia 1 commissioning phases with innovative solutions in data processing and archiving. INPE is in discussions with China for the next CBERS satellite missions.

INPE has a connector to CWIC developed by the Connected Assets initiative, and is now moving to OpenSearch service. Lubia mentioned having problems with definitions of collections and granules for OpenSearch and for IDN; she requested more discussion in WGISS on this topic.

INPE is successfully monitoring the Amazon biome and sees the need to do the same with the other five biomes in Brazil, especially for both deforestation and land use cover. INPE is working to improve dissemination of PRODES, and has signed a cooperation agreement with EU in order to make INPE a regional hub in Latin America for the dissemination of Copernicus data. The specifics of the implementation have not been discussed yet.

## JAXA

Makoto Natsuisaka gave a report on the Japan Aerospace Exploration Agency (JAXA). He began with an overview of the organization, highlighting the EO-related groups. He also displayed the location of their offices, facilities, and centres, and listed past, current and future missions.

Makoto discussed the GCOM-C mission for climate, launched at the end of 2017, explaining the main targets. He showed images showing forest change, snow, sea ice, water and ice cloud, and volcanic ash.

Makoto described JAXA’s ground segments, and improvements to the G-Portal, which is the primary gateway for JAXA global EO standard products. He also describe the portals for environmental products and for disaster prevention, and JAXA’s long-term plan for satellite missions.

## NASA

Chris Lynnes gave a presentation on the NASA Earth Observing System Data & Information System (EOSDIS), discussing its role and the scope of its data collection. He described their data sources, data centres, and data access. Chris noted that data access is centralized with reusable capabilities, and that the Earthdata search tool has been released, as have the Global Imagery Browse Services, and Worldview.

Chris discussed significant progress and accomplishments of EOSDIS, and data volume statistics.

## UKSA

Esther Conway gave a report on the UK Space Agency (UKSA). She described the agency’s responsibilities and vision for data access and exploitation. She noted that UKSA has a fully operational Collaborative Ground Segment. Esther described the UK activities for data preservation and stewardship, and for data access, formats, and interoperability. UKSA is also active in studying quality indicators in discovery metadata and uncertainty, and in standardisation and best practices.

Esther listed and described the FDA platforms and elements landscape for platforms, data cubes, and analysis ready data. She concluded saying that they are hosting the PV 2018 Conference in May.

Makoto asked if UKSA works in the areas of disasters and agriculture; Esther replied that the agency is newly formed so discussion is underway on these topics. Her role is to report back to UKSA about the activities of CEOS and WGISS.

Michael asked if the UK data collections will be available through FedEO; this will be investigated.

**Action WGISS-45-15:** Esther Conway, Andrea Della Vecchia, and Michael Morahan to ensure that UKSA’s data collections are accessible through FedEO.

# WGISS Plenary Part II

## Future Meetings

Robert Woodcock gave a presentation on the next WGISS meeting (WGISS-46), which will be hosted by DLR in Oberpfaffenhofen, Germany, the week of October 22, 2018. The meeting will likely be four days, Monday to Thursday, and work on the agenda will begin soon. Robert provided logistics and accommodation options.

## WGISS Summary

Mirko Albani gave a summary of the meeting. He began by summarizing the CEOS Chair objectives and priorities. He focused on WGISS cooperation with CEOS and GEO:

* CEOS Carbon Team with ongoing development of prototype Carbon Portal.
* WGDisasters opportunities with Recovery Observatory (RO), Disaster Risk Management (DRM) Demonstrators
* WGClimate with Ongoing: review of ECV inventory to ensure CDRs access through WGISS Connected Data Assets Infrastructure
* WGCapD Ongoing: WGISS Technology webinars publicized through WGCapD. Opportunities: promote awareness on FDA through WGCapD (FDA-5), storing WGCapD training material / documentation on WGISS Web site, explore joint discovery of data and associated information
* WGCV Ongoing: Several addressed during Joint Session
* SEO Ongoing: COVE tool harvesting metadata from WGISS Connected Data Assets Infrastructure. Opportunities: jointly develop approach for on demand data-cube creation
* Virtual Constellations. Opportunities: will be addressed at SIT-33
* GEO. Ongoing: WGISS contributing to GEOSS-EVOLVE initiative in the GEO 2017-19 Work Plan. Ongoing: participation & support to GEOSS Data Providers Workshops (e.g. “User Metrics” session at upcoming 3rd GEOSS Data Providers Workshop).

Mirko discussed the development of the WGISS Work Plan 2018-20 which is underway, and highlighted the CEOS Work Plan 2018-20, with contact names included in each of the actions that pertain to WGISS.

Mirko mentioned the need for agency or SEO support for the WISP, since NOAA will no longer be able to support it.

WGISS plans:

* To create a web page on best practice and white papers, covering publications for all working groups.
* To draft a white paper for a process to alert to the availability of new dataset discovery and access.
* To create a White Paper on Data Stewardship Reference Model (not more than 2 pages).
* To update Data Preservation Guidelines and Preservation Workflow vs DMP, WGISS Data Management and Stewardship Maturity Matrix
* To provide a proposal to GEOSS-EVOLVE and DMP Team

Mirko summarized the activities of the WGISS Connected Data Assets team, including the CWIC, IDN, FedEO, and Carbon Portal. He also listed the recent technology webinars, and reminded of the discussions on single sign-on and user metrics, OSS inventory, OGC Testbed 13/14, and OpenSearch advances.

The joint session with WGCV included four broad topics; the resulting tasks are:

1. Data Formats and Interoperability in the framework of FDA. Work to define a set of quality/uncertainty parameters to include within ARD and determine an initial method to incorporate these along with their provenance into the data stream, including education of the data community on how to provide provenance information.
2. Quality Indicators in Discovery Metadata. QI Test case. Evaluate whether SST would provide a suitable test case for QI development that can act as an expressive case for QI access and implement this or an alternate case as determined through this activity
3. CEOS Data Cubes and CEOS Test Sites Data Access in support of WGCV Activities. LPV supersite data access test case: LPV supersites are a good test case related to validation with a broad user community. Objective is to provide an improved user access to the LPV supersite data sets coupled with analytics tools to provide improved results for the users
4. Standardization and Best Practices (e.g. ISO 19159-3). Revisit and update WGISS-41/WGCV-40 actions: Define a point of contact within WGCV to work with WGISS to reinvigorate the actions of WGISS-41 in terms of more recent CEOS activities and move towards closure of these items to provide a good practice white paper for inclusion on the CEOS website

The FDA Workshop included sessions on FDA functional blocks, data cubes, dynamic gridding, DIAS, Thematic Exploitation Platforms, mission application platforms, agency cloud applications, cloud hosting, and cloud analytics.

## WGISS-45 Actions

The following actions resulted from the meeting.

**Action WGISS-45-1:** Call for interest from a WGISS-participating agency to take over WISP: deadline for expression of interest is end of April 2018. WGISS Chair will ask SEO to support these services in case no other agency actively involved in WGISS volunteers to take over WISP.

**Action WGISS-45-2:** ESA to update the draft WGISS Work Plan with the results of WGISS-45 session discussion; final draft to WGISS-Exec for review by end of April; issue by end May 2018.

**Action WGISS-45-3:** Michelle Piepgrass to organize a telecon in May (WGISS, WGCapD, SEO, CEO) to identify ways for WGISS to support WGCapD on CEOS Workplan action FDA-5.

**Action WGISS-45-4:** Michelle Piepgrass to provide Interest Group leads the WGISS brochure (soft copy) presented at WGISS-44; Interest Group leads to review and simplify their text, keeping the mission of WGISS by April 25. WGISS Exec to consolidate and finalize the brochure, and create a page on the WGISS website for easy access.

**Action WGISS-45-5:** Mirko Albani and Iolanda Maggio to organize development of a procedure for making visible to CEOS the new agency datasets that become available by Q4 2018.

**Action WGISS-45-6:** DSIG to create a White Paper on the Data Stewardship Reference Model and Maturity Matrix and send it to the DMP team as a proposal for inclusion in the GEO DMP by July 2018.

**Action WGISS-45-7:** Yonsook Enloe, Richard Moreno, Rob Woodcock, Mirko Albani, and Chris Lynnes to identify and submit to Michelle the valid versions of their Best Practice and White Papers. Michelle to display these on the Document Management Table on the WGISS website by end of June 2018.

**Action WGISS-45-8:** Mirko Albani, Iolanda Maggio and Michelle Piepgrass to address Best Practice and White Papers (not lessons learned) Web page (to be put at highest level in CEOS Web Site) population with other WGs material after completion of action on WGISS Interest Groups. Maintain and populate page by September 2018.

**Action WGISS-45-9:** Liping Di to send an email to Mirko Albani with the details of information he still needs from the Carbon team. Mirko to forward these to the CEOS Carbon Team and organize a teleconference to review the status of the portal so that it can be opened to the broader community for review by April 20, 2018.

**Action WGISS-45-10:** Iolanda Maggio to obtain from Robert Woodcock the CSIRO information model for managing OSS. Iolanda to distribute revised software metadata model to agencies, asking them to complete the relevant information and return to Iolanda by mid-May.

**Action WGISS-45-11:** Mirko Albani and Iolanda Maggio to ask every agency to send information on how they collect user metrics, which metrics they collect, and what they do with their metrics by end of April.

**Action WGISS-45-12:** Mirko Albani to work with Paola De Salvo to organize a GEO-CEOS interoperability session at WGISS-46 in order to identify how WGISS can cooperate with the interoperability goals of GEOSS by end of May, 2018.

**Action WGISS-45-13:** Michael Morahan to follow up with Greg Stensaas to update the test site information in the IDN by end of June.

**Action WGISS-45-14:** WGISS Exec to organize bi-monthly telecon with WGCV to formalize as actions the tasks identified during the Joint Session, and to review status of these. First telecon in June.

**Action WGISS-45-15:** Esther Conway, Andrea Della Vecchia, and Michael Morahan to ensure that UKSA’s data collections are accessible through FedEO.

## Concluding Remarks

Mirko Albani concluded the meeting thanking INPE for excellent hosting; logistics, facilities, and activities were fabulous. Mirko also thanked the participants for their contributions toward an excellent meeting.

.

# Glossary of Acronyms

API Application Programming Interface

ARD Analysis Ready Data

AWS Amazon Web Services

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

COTS Commercial Off-the-Shelf

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

DC data cube

DIF Directory Interchange Format

DOI Digital Object Identifier

ECV Essential Climate Variable

EO Earth Observation

ESIP Federation of Earth Science Information Partners

GCI GEOSS Common Infrastructure

GCMD Global Change Master Directory

GEO Group on Earth Observations

GEO-GLAM Global Agricultural Monitoring

GEOSS Global Earth Observation System of Systems

GFOI Global Forest Observations Initiative

GHG Greenhouse Gas

GIS Geospatial Information System

GPM Global Precipitation Mission

GPU Graphics Processing Unit

GSDI Global Spatial Data Infrastructure

GUI Graphical User Interface

HPC High Performance Computing

ICT Information and Communication Technology

IDN International Directory Network

ISO International Standards Organization

LSI Land Surface Imaging

LTO Linear Tape-Open

MOU Memorandum of Understanding

NRT Near real-time

NWIP New Work Item Proposal

OGC Open Geospatial Consortium

PI Persistent Identifier

POC Point of Contact

RSS Rich Site Summary

SEO Systems Engineering Office

SDCG Space Data Coordination Group

SIT Strategic Implementation Team

SLT System Level Team

SWG Standards Working Group.

TOA Top of the Atmosphere

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