MINUTES

OF THE

42nd MEETING

OF THE

CEOS WORKING GROUP ON
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Frascati, Italy

19 September to 22 September, 2016

Hosted by

European Space Agency

(ESA)

Table of Contents

[1 WGISS Plenary Session, Part I 5](#_Toc465180654)

[1.1 Chair Welcome, Introductions, Adoption of Agenda 5](#_Toc465180655)

[1.2 Host Welcome and Logistics Information 5](#_Toc465180656)

[1.3 ESA Opening Address 5](#_Toc465180657)

[1.4 WISP Report 5](#_Toc465180658)

[1.5 WGISS Chair Report 6](#_Toc465180659)

[1.6 Review of WGISS-41 Actions 6](#_Toc465180660)

[1.7 GEO and GEOSS Report 7](#_Toc465180661)

[1.8 SEO Report 7](#_Toc465180662)

[1.9 Review of CEOS and GEO Actions 8](#_Toc465180663)

[2 Agency and Liaison Reports 10](#_Toc465180664)

[2.1 US Geological Survey (USGS) 10](#_Toc465180665)

[2.2 National Aeronautics and Space Administration (NASA) 10](#_Toc465180666)

[2.3 Japan Aerospace Exploration Agency (JAXA) 11](#_Toc465180667)

[2.4 Indian Space Research Organization (ISRO) 11](#_Toc465180668)

[2.5 Global Spatial Data Infrastructure (GSDI) Association 11](#_Toc465180669)

[2.6 National Oceanic and Atmospheric Administration (NOAA) 12](#_Toc465180670)

[2.7 Geoscience Australia (GA) 12](#_Toc465180671)

[2.8 National  Institute for Space Research (INPE) 12](#_Toc465180672)

[3 Data Use 13](#_Toc465180673)

[3.1 Future Data Architectures 13](#_Toc465180674)

[3.2 Data Cube Projects at ESA 14](#_Toc465180675)

[3.3 Open Source Big Earth Observation Data Analytics at INPE 14](#_Toc465180676)

[3.4 Use Discussion 15](#_Toc465180677)

[4 Data Access 16](#_Toc465180678)

[4.1 International Directory Network (IDN) 16](#_Toc465180679)

[4.2 CEOS OpenSearch II Project 16](#_Toc465180680)

[4.3 OpenSearch for EO Evolution 17](#_Toc465180681)

[4.4 Federated Earth Observation (FedEO) 17](#_Toc465180682)

[4.5 CEOS WGISS Integrated Catalog (CWIC) 17](#_Toc465180683)

[4.5.1 CWIC Report 17](#_Toc465180684)

[4.5.2 EUMETSAT Report 18](#_Toc465180685)

[4.5.3 ISRO Report 18](#_Toc465180686)

[4.5.4 INPE Report 18](#_Toc465180687)

[4.5.5 NOAA Report 19](#_Toc465180688)

[4.6 WGISS Connected Data Assets 19](#_Toc465180689)

[4.7 Unified Metadata Model: WIGOS/CGMS Mapping 20](#_Toc465180690)

[4.8 Access Discussion 21](#_Toc465180691)

[5 Data Preservation 22](#_Toc465180692)

[5.1 Data Stewardship Interest Group Update 22](#_Toc465180693)

[5.2 Preservation of Associated Knowledge Best Practices 22](#_Toc465180694)

[5.3 Maturity Matrix/Model for Harmonization 22](#_Toc465180695)

[5.4 Report on Agency Stewardship Activities 23](#_Toc465180696)

[5.4.1 ESA - Heritage Data and Knowledge Preservation 23](#_Toc465180697)

[5.4.2 NASA - Earth Science Data Stewardship 23](#_Toc465180698)

[5.5 Data Preservation Discussion 24](#_Toc465180699)

[6 Technology Exploration Workshop on CLOUD COMPUTING 26](#_Toc465180700)

[6.1 Introduction 26](#_Toc465180701)

[6.2 GS Evolution and EO Innovation Europe 26](#_Toc465180702)

[6.3 ESA Thematic Exploitation Platforms and Cloud Computing Activities 26](#_Toc465180703)

[6.4 Cloud Computing and Security 26](#_Toc465180704)

[6.5 JAXA Approach on Virtualization and Cloud Computing 27](#_Toc465180705)

[6.6 ISRO Requirements and Research Issues for EO Data Processing Cloud 27](#_Toc465180706)

[6.7 Computing in the Cloud at Geoscience Australia 28](#_Toc465180707)

[6.8 Leveraging the Value of Data with Industry at NOAA 28](#_Toc465180708)

[6.9 Cloud Hosting at USGS 28](#_Toc465180709)

[6.10 Assessing Applications of Cloud Computing to NASA’s EOSDIS 29](#_Toc465180710)

[6.11 Data Cube Use of Cloud Computing by CEOS 29](#_Toc465180711)

[6.12 Cloud Computing Discussion 29](#_Toc465180712)

[7 WGISS Plenary, Part II 31](#_Toc465180713)

[7.1 Future Webinar Discussion 31](#_Toc465180714)

[7.2 Future Meetings 31](#_Toc465180715)

[7.3 Chair Summary 31](#_Toc465180716)

[7.4 WGISS-42 Actions 32](#_Toc465180717)

[7.5 Adjourn 34](#_Toc465180718)

[8 Glossary of Acronyms 35](#_Toc465180719)

List of Participants

ASI Luigi Mascolo

CAS Lizhe Wang, Jining Yan

CCMEO Jeff Cote\*

CNES Richard Moreno, Julien Airaud, Benoît Chausserie Laprée, Jérôme Gasperi

CSIRO Robert Woodcock\*

DLR Katrin Molch, Stephan Schropp

ESA Mirko Albani (WGISS vice-Chair), Olivier Barois, Philippe Bally, Yves Coene, Andrea Della Vecchia, Damiano Guerrucci, Guenther Landgraf, Henri Laur, Sveinung Loekken, Cristiano Lopes, Iolanda Maggio, Philippe Mougnaud, Salvatore Pinto

EUMETSAT Uwe Voges, Harald Rothfuss

GEO Secretariat Osamu Ochiai\*

Geoscience Australia Simon Oliver

GSDI/HUNAGI Gábor Remetey-Fülöpp

INPE Lubia Vinhas

ISRO Nitant Dube

JAXA Satoko Miura, Masumi Matsunaga, Shinichi Sekioka

NASA Andrew Mitchell (WGISS Chair), Dawn Lowe, Simon Cantrell\*, Eunice Eng\*, Yonsook Enloe\*, Lingjun Kang\*, Brian Killough\* (CEOS-SEO), Chris Lynnes\*, Mark McInerney, Michael Morahan, Doug Newman\*, Eugene Yu\*, Michelle Piepgrass (WGISS Secretary)

NOAA Martin Yapur, Anne Kennerley, Ken McDonald

NRSC Sai Kalpana Tanguturu\*

ROSCOSMOS Tamara Ganina\*

UKSA Christopher Hall\*

USGS Kristi Kline

\* Via web conference or email

# WGISS Plenary Session, Part I

## Chair Welcome, Introductions, Adoption of Agenda

Andrew (Andy) Mitchell (WGISS Chair) welcomed the participants to WGISS-42. Andy thanked ESA-ESRIN for all the excellent arrangements, and asked those present to introduce themselves.

Andy noted that the meeting has a full agenda, and everyone is looking forward to the Cloud Computing Workshop. He reviewed the agenda and it was adopted with no modifications.

## Host Welcome and Logistics Information

Mirko Albani welcomed the participants to the meeting. He described the facility, lunch, breaks, Wi-Fi, and contents of welcome package. He also described the events of the week: tour of Rome, and tour of ESRIN facility, including the ESRIN Control Room for EO Payloads and the Astronomical Observatory.

## ESA Opening Address

Henri Laur, ESA Earth Observation Missions Management Office, ESA, welcomed the participants on behalf of ESA and ESRIN. He stated that the purpose of ESA is to provide for and promote, for exclusively peaceful purposes, cooperation among European states in space research and technology and their space applications. ESA is one of the few space agencies in the world to combine responsibility in nearly all areas of space activity. ESA has 22 Member States: 20 states of the EU (AT, BE, CZ, DE, DK, EE, ES, FI, FR, IT, GR, HU, IE, LU, NL, PT, PL, RO, SE, and UK) plus Norway and Switzerland. Seven other EU states have Cooperation Agreements with ESA: Bulgaria, Cyprus, Latvia, Lithuania, Malta, Slovakia and Slovenia. Discussions are ongoing with Croatia. Canada takes part in some programmes under a long-standing Cooperation Agreement.

Henri described ESA’s budget, noting that in addition to the member state contributions (principal donors Germany and France), a good portion of the income comes from third party donors. He also described the budget allocation by domain, where EO has the largest percentage.

The EC and ESA share the common aim to strengthen Europe and benefit its citizens. Closer ties and an increased cooperation between ESA and the EU bring substantial benefits to Europe by guaranteeing Europe’s full and unrestricted access to services provided by space systems for its policies. The relationship also helps with encouraging the increasing use of space data to improve the lives of its citizens, increasing political visibility of space and taking full benefit from its economic and societal dimension. Space science is perceived as having a major impact on its citizens.

As a European research and development organisation, ESA is a programmatically driven organisation, i.e. the international cooperation is driven by programmatic needs and rationale. ESA has strategic partnerships with USA, Russia and China, and long-standing cooperation with Japan, India, Argentina, Brazil, Israel, South Korea, Australia and many more. With EU members that are not ESA member states there is enhanced cooperation and joint activities.

About 85% of ESA’s budget is spent on contracts with European industry. The industrial policy ensures that Member States get a fair return on their investment, improves competitiveness of European industry, maintains and develops space technology, and exploits the advantages of free competitive bidding, except where incompatible with objectives of the industrial policy.

ESRIN is ESA’s centre for EO, where operations and exploitation of EO satellites are managed. It is a facility with 200 staff and 300 contractors active in Earth Observation, Vega Department, Corporate Informatics, Telecommunications, Contracts, Site, Personnel, Communication, and ESA Security Office. ESRIN is the management centre for the VEGA Small Launcher Programme. VEGA is able to place 2500 kg satellites into polar and low-Earth orbits, and has had seven successful launches.

The ESA Ministerial Council will meet in December 2016; at that time several optional programmes will be described. In addition, two key elements in the ESA General Budget 2017-21 will be discussed: the need for supporting heritage data, and Earthnet, the European International Gateway for EO.

ESA’s EO missions are grouped in three categories: meteorological missions, Copernicus missions, and science missions. Future missions are grouped in four blocks: industrial studies, mission development, mission management, and science for society. In these blocks are included the Copernicus Evolution Instrument Models and the Earth Explorer missions exploitation phase.

Henri described the Copernicus data access and redistribution, with Sentinel data hubs operated by ESA. The Copernicus Ground Segment features dedicated data access infrastructure solutions tailored to the needs of the various use topologies: large and small private companies, access to anyone through data hub, collaborative mirror sites, international partners mirror sites, and provision of higher level products.

The Copernicus Space Component (CSC) Ground Segment architecture implements the above policy, and includes an evolutionary approach to further enhance the data exploitation by the broad user community. The CSC Space and Ground Segment evolution benefits from the innovation activities funded through ESA programmes (e.g. EO Envelop Programme).

Useful links: [www.esa.int](http://www.esa.int), earth.esa.int, sentinels.copernicus.eu

Andy commented that a group like WGISS is here to help with having one voice on the topic of heritage data. WGISS develops white papers, reports, and recommendations that go to higher levels to add value.

## WISP Report

Anne Kennerley presented the WGISS Infrastructure Support Project, whose goal is to support WGISS in its activities. The lead of WISP is Martin Yapur, supported by Anne Kennerley, Kim Holloway, and Michelle Piepgrass.

WISP has been active in revamping of WGISS webpages based on WGISS functions and continual upkeep of the webpages and mailing lists. The WISP’s success depends on members providing mailing list and web page updates, and seeking support for outreach activities.

Anne gave instructions for submitting presentations and for remote participation.

**Action WGISS-42-25:** WISP team to compile a mailing list of members who regularly attend WGISS meetings for specific communications.

## WGISS Chair Report

Andy reported on recent activities of WGISS and CEOS issues related to WGISS. He began with a discussion of the recent SIT Technical Workshop, where partnerships with development banks, the UN, data giants, and NGOs were discussed. CSIRO identified this topic for further discussion at the CEOS Plenary, stressing that many user groups are unaware of benefits from using EO data.

Reports on the CEOS data cube and ARD are that the first FDA-related pilots are emerging bottom-up within CEOS; work is well advanced within SEO/SDCG/LSI-VC. The CEOS Data Cube Work Plan has been developed to document and communicate the many threads and dimensions, and the ARD specification is progressing well. The 3-year CEOS Data Cube Work Plan and the CARD4L high level definition have both been endorsed.

The Future Data Architectures (FDA) short-term recommendations were well received, and the SIT will recommend to the Plenary to endorse the 1-year extension of the FDA team; continued relevant work by WGISS was encouraged. The team intends to propose to Plenary an integration of existing CEOS activities as a more focused demonstration of FDA benefits for a small-scale ARD production in a CEOS data cube framework and the GFOI Colombia demonstrator.

CEOS is holding several carbon actions, so the focus will be on five to seven VC/WG initiatives:

* ACC GHG Constellation Paper
* WGClimate Carbon gap analysis (ECVI)
* WGISS portal
* WGCV actions
* NASA biomass calibration/validation and products
* JAXA IPCC TFI engagement

Plenary will need to agree on the overall approach and comment on individual initiatives; CEOS will continue with GEO Carbon Initiative engagement, mapping agency level projects onto Carbon actions, and plan for a 2-yearly CEOS Carbon Workshop.

At the VCs/WGs side meeting Andy asked for input on identifying the products these groups have difficulty discovering or accessing, and if they have a need for science discipline-focused data access portals. He also asked if there are technologies that WGISS should prioritize in the Technology Exploration webinars. The groups took an action to reply to these questions. The feedback received at the meeting was: Can WGISS allow for the discovery of data services available with connected data assets; can WGISS track user metrics on connected data assets; can WGISS work on including in-situ in their interoperability efforts.

It was agreed that the next step is for WGISS to search the MIM for targeting CEOS data that are not available in IDN nor available via WGISS standards.

Participants were asked to complete the CEOS information systems survey by end of the September: <https://www.surveymonkey.com/r/ceos-info-systems>

Andy noted that the Coordination Group for Meteorological Satellites (CGMS) has a working group similar to WGISS: CGMS WG IV provides a regular forum for CGMS agencies to address topics of interest in areas related to data access in general and the contribution to the WMO Information System (WIS). The Working Group addresses issues related to data dissemination systems, data formats and metadata exchange, and it also deals with the user interfaces and data access. Andy targeted these WG IV actions for WGISS participation:

A43.06: CGMS members to provide a listing of their data access portals.

A44.01: To submit the “Guidance Documentation on WMO Core Profile Metadata Creation for Satellite Products” to WMO IPET-MDRD and IPET-SUP.

A44.03: CGMS members (data providers) to

a) discuss and respond to the recommendation from CGMS-44-CEOS-WP-02: CEOS recommends the adoption of the WGISS supported standards for searching Climate Data Records (CDRs). WGISS will provide technical support to CGMS data providers providing their climate data records through the WGISS data access infrastructure (IDN, CWIC, FedEO); and

b) report how far the standards WGISS developed (as described in CGMS-44-CEOS-WP-02) are supported.

R42.01: Satellite operators to provide WIS Discovery Metadata Records, compliant to WIS requirements and following the guidance to be provided by the CGMS-WMO Task Force on metadata implementation, in order to facilitate satellite information discovery and access

Richard asked if the FDA document/report has another year to evolve; Andy said yes; there has been a lot of feedback on the current document; for example, how to handle what is currently in the appendix, and what should the recommendations be. Richard concurred that there is still a lot more to do; the document will be a living document.

Nitant commented that CGMS agencies are the same as in WGISS, and an effort should be made to harmonize the activities of both groups. Martin recommended that the CGMS representatives be invited to the WGISS-43 meeting; Nitant suggested identifying specific people.

Satoko asked about the convergence of the FDA and RDA topics. Andy replied that originally FDA and ARD were together, but now are being dealt with as two separate topics.

Mirko commented on the issue of exploring the access to the services: in GEO they are trying to access services, and perhaps CEOS can draw from them. Nitant added that people are more interested in services now, rather than data; services are more mature, so it makes sense to start working on this.

## Review of WGISS-41 Actions

Andy reviewed the actions from WGISS-41, discussing the outcomes of each.

## GEO and GEOSS Report

Osamu Ochiai\* presented a report for the GEO Secretariat. He noted that this is a unique time in GEO’s history, with a transition to the next decade 2016-2025, and recognition of GEO’s convening power of members, POs, development banks, foundations, and the emerging commercial sector. He also noted the evolution and recognition of policy mandates, and a new Strategic Plan with new programmatic mechanisms – community activities, foundational tasks, initiatives and flagships.

The governance of GEO is through a Plenary with 103 Member Countries, 103 Participating Organizations and 12 Observers; a GEO Executive Committee, a GEO Programme Board, and the GEO Secretariat. There are four types of GEO implementation mechanisms: GEO Community Activities, GEO Initiatives and GEO Flagships, all of which interact with GEO Foundational Tasks.

From the current list of Foundational tasks, Osamu noted two that are pertinent to WGISS; Osamu agreed to forward the requirements document to Andy for WGISS review:

GD-02 GCI Operations (including access to knowledge) will remain the Foundational Task. Perform GCI Components operations (GEOSS Portal, GEO DAB, and Registries). Maintain partnership with Data and Service Providers and improve these Providers discoverable and accessible. Connect new providers which are relevant to Flagships and key members and participating organizations. Collect requirements and feedback from User Communities and Stakeholders.

GD-07 GCI Development (includes development of data management guidelines) will likely be moved to an Initiative (still in discussion; decision will be made in the next month). This will develop a GEOSS Architecture based on documented and emerging user requirements, develop and test new GCI functionalities, solutions, and components, develop a process to implement the Data Management Principles Guidelines for providers; promote the advancement of GEOSS interoperability through the Standards and Interoperability Forum (SIF); develop the Community Portal Recommendations.

Osamu reminded WGISS that the GEOSS implementation requires the application of the Data Sharing Principles, meaning full and open exchange of data and products at minimum time delay and cost, and free of charge or cost of reproduction. The Data Management Principles Strategy is based on the notion that the value of Earth observations are maximized through data life-cycle management based on ten principles supporting five themes: Discoverability, Accessibility, Usability, Preservation, and Curation.

Osamu displayed a diagram showing data in GEO-DAB, with a large proportion coming from CEOS. Potential issues that have been identified are data accessibility mainly in the FedEO catalog, and the need for better metadata. Osamu proposed that CEOS WGISS to identify a point-of-contact to discuss further improvements to the discoverability and accessibility (through IDN, CWIC and FedEO) of CEOS agency assets.

Osamu also mentioned the Data Providers’ side event in the GEO-XII Plenary in November, with objectives of establishing a two-way dialogue with data providers to improve the discoverability, accessibility and usability of GEOSS resources. Data providers already contributing to GEOSS and new data providers, flagships and initiatives and users are invited and encouraged to participate to ensure that the key objectives of the workshop are met. This will help the GEO community define priorities and shape the agenda of a more comprehensive event to be held in early 2017.

Mirko said that he would contact Osamu immediately on the FedEO accessibility issue, and noted that WGISS is discussing a connected data assets system-level team to coordinate with GEO.

Mirko asked what is meant by options and procedures for certification of data providers. Osamu replied that this has to do with standards on data quality and citation, and how the process can be applied in the DMPs. Mirko asked if GEO is planning on adopting a certification method for its data providers. Osamu said this is a sensitive topic, and is only in discussion. Andy asked if the data providers’ side event has virtual meeting capability; Osamu replied that they plan a WebEx, and are in discussion with the host about bandwidth.

Richard asked what the consequence is of moving the GD-07 to an “Initiative”, and if it is, is it possible to move the DMP portion to GD-02. Osamu indicated that it may not be possible to coordinate the request. Richard emphasized the importance of DMP, and Osamu said he could send an email about this, and would copy Mirko and Andy.

**Action WGISS-42-01:** Andy Mitchell and Mirko Albani to recommend to the GEO-Sec (Osamu Ochiai), if GD-07 becomes an initiative in the GEO Work Plan 2016-18, to move the Data Management Guidelines Task to GD-02.

## SEO Report

Brian Killough reported for the CEOS Systems Engineering Office (SEO). He began with a reminder of the CEOS Visualization Environment (COVE) tool, from which they are getting a lot of positive feedback. Recently they have moved from Google Earth to Cesium globe interface with a global phenology overlay. Future plans are to improve the coverage analyzer, and add more links to mission archives.

Brian reported on the following topics:

* Data cubes – The SEO is leading an effort in CEOS to develop an open source data cube architecture for data management and enhanced applications. This is part of the CEOS “Future Data Architectures” effort.
* Analysis Ready Data – The SEO is working with the Lands Surface Imaging Virtual Constellation (LSI-VC) team to develop a description and technical specification for CEOS Analysis Ready Data for Land (CARD4L). Draft documents are available from the CEOS SIT Workshop last week.
* Data Interoperability – There is a strong desire to use the data cube architecture to test data interoperability options. Two cases are combining optical and SAR, or using optical datasets with different resolutions.
* Cloud Computing – The SEO is investigating cloud-based computing options with Amazon Web Services (AWS) to support data cubes.

Brian also gave a brief description of data cubes, noting that their unique feature is to exploit time series and increase data interoperability. He described the CEOS data cube, working with CEOS space agencies to develop plans for sustained provision of Analysis Ready Data (ARD). It is open source software, developed and sustained by CEOS, with support for diverse datasets, and deployment via local computers, regional hubs (e.g. SERVIR), or computing cloud (e.g. Amazon), and connections to common GIS tools (e.g. ArcGIS, QGIS) and Advanced Programming Interfaces (APIs) for users.

The data cube work plan provides a reference for internal and external data cube activities as there is great interest in data cubes and Future Data Architectures (FDA). The majority of the work is managed and funded by the SEO with significant contributions by CSIRO and GA. The SEO works closely with Australia to utilize elements of the AGDC development and communicates with USGS regarding its plans for LCMAP. The document captures expected outcomes, task descriptions and target dates of completion in the areas of core software (ingestors, GIS tools, GUI tools), data preparation (ARD), user engagement (GFOI, GEOGLAM), capacity building (World Bank, WGCapD), and prototypes.

CEOS is developing the following data cube prototypes:

* Colombia – The government (IDEAM) and Andes University teams have made considerable progress in learning how to create and use data cubes. Land change detection and water detection are the primary application needs. Future plans will add many more datasets and applications.
* Kenya – Recent changes in the government have caused uncertainty in the plans for a data cube project in 2017. Australia and Clinton Foundation have terminated their work.
* Lake Chad, Africa – Considerable interest from World Bank in using a data cube for time series analysis of land and water in the Lake Chad region. Possible project to begin in mid-2017, pending approval.
* Asia Mekong – Investigating possible project with SERVIR and JAXA to serve data cubes to the Mekong region.
* Balkans – Recent proposal submitted to World Bank to develop a data cube to support multiple applications in Albania. Proposed start by mid-2017.
* Switzerland – SEO approached by UNEP GRID Geneva and the University of Geneva to develop a data cube pilot project. Significant computing and programming resources exist, so little effort is needed to get them started.
* Disasters Pilot – Recent discussion with David Green (NASA Disasters Lead); evaluating the potential to test the
SLIP-DRIP landslide analysis code with a data cube.

Brian requested WGISS support to continue to expand the connections from mission archives to the COVE tool (future targets: Sentinel-2 and CBERS-4). He also asked for assistance to find an approach for automated discovery, processing, downloading and ingesting of data to support users with data cubes.

Martin asked if they have a machine-to-machine interface; Brian said that they do for Landsat and are working toward the same with Sentinel. Lubia added that for CBERS-4 INPE has the tool, an API being used by CWIC. Mirko asked why they are getting Sentinel from Mirror sites. The reason is better working connection; with distance robustness can be lost so mirror archives provide faster access.

## Review of CEOS and GEO Actions

Andy Mitchell presented the tasks from the CEOS Work Plan 2016-18 that are pertinent to WGISS:

DATA-2: Full representation of CEOS Agency datasets in the IDN and accessible via WGISS Interoperable Standards. WGISS began discussions with ISRO and the following Australian centers in order to get their data accessible via WGISS interoperable standards (i.e. IDN, CWIC). (Geosciences Australia / CSIRO / Bureau of Meteorology /Australian National University &National Computational Infrastructure). New entries were added to the IDN from ESA, EUMETSAT, and JAXA datasets.

VC-1:  List of Relevant Datasets from VCs. WGISS is requesting updated list from the VCs.

VC-25: Increase the visibility of land surface imaging data holdings. WGISS will work in conjunction with the LSI-VC to ensure relevant datasets are visible through WGISS Interoperable Standards.

OUT-7: Data cube infrastructure development. WGISS will support the SEO in the development of a general CEOS data cube infrastructure

CARB-9: GFOI Data Services Pilot Projects for Kenya and Colombia. WGISS will support the SEO in the delivery of Data Services Pilot Projects (based on the data cube architecture) for Kenya and Colombia.

CARB-08-35: The CEOS Carbon Subgroup (recommended in Carbon-Action-38) will develop guidelines for appropriate data use of satellite data and data products. This will require improved interaction between the carbon cycle community and the satellite community; comprehensive review of the current use of data products, including current data limitations; and reconciliation of methodological differences and spatial compatibility. Such interactions may include co-sponsorship of joint workshops targeting specific data needs and investment in community product assessments, especially for key in comparison exercises.

Two WGSS initiatives in support of the carbon actions were presented at the SIT Technical Workshop, 2016:

* For ECVs/CDRs Discovery and Access though WGISS Systems, whose objective is to facilitate discoverability and accessibility of ECV products and space-borne CDRs relevant for the CEOS Carbon Action via WGISS Interoperability Systems and Standards. The approach is to start from results of the WGClimate Questionnaire for ECV Inventory population, tailoring for Carbon Action and gaps identification with respect to data records already discoverable/accessible through WGISS systems (Q1/Q2 2017). This will be followed by feasibility analysis, priorities setting and liaising with relevant organizations (Q2/Q3 2017), and start of technical activities (Q3/Q4 2017). Support from WGClimate and experts will be necessary for setting priorities; activities at data providers’ side to be carried out by relevant entities.
* For the WGISS Carbon Portal, whose objective is the development of a CEOS WGISS Carbon Portal prototype similar to the Water portal (<http://waterportal.ceos.org/>) to allow displaying Carbon datasets and providing assistance to scientists and general users in the development of related services and tools. The approach is the collection of needs from Carbon (or WGClimate) experts on what needs to be in the portal (Q1/Q2 2017); Carbon Portal requirements definition and system design (Q2/Q3 2017); start development (Q3/Q4 2017). Requirements definition and development resources will be provided by NOAA.

For the 2016 GEO Work Program, WGISS has been added as a contributor to the Foundational Task GD-7 Subtask 2 ‘GCI Development’ and GD-2 ‘GCI Operations’. As a contributor, WGISS will continue to advocate for CEOS agency mission data to be contributed to the implementation of the GEOSS via the WGISS interoperable systems and standards (e.g. IDN, CWIC, and FedEO). As more CEOS data providers adopt WGISS supported standards (e.g. OGC CSW 2.0.2 and CEOS OpenSearch Best Practices) these data providers holdings will be made discoverable via these systems. WGISS will also maintain the website titled ‘Connected data assets’ to provide up-to-date metrics for IDN, CWIC and FedEO. WGISS will contribute to this activity through the WGISS Technology Exploration Interest Group which serves as a forum for the exchange of technical information and lessons-learned about current and trending software technologies, services, and other internet-based software technologies.

Andy also described updates to the GEOSS Portal, and announced the Virtual Workshop GEO DAB API’s at the end of September.

**Action WGISS-42-02:** Andy Mitchell and Mirko Albani to obtain from WGClimate the final results of the ECV Inventory Gap Analysis for Carbon.

**Action WGISS-42-03:** Ken McDonald to research the GEO Carbon Portal.

**Action WGISS-42-04:** Andy Mitchell, Mirko Albani, Martin Yapur, and Ken McDonald to define the requirements for a CEOS Carbon Portal, working with WGClimate and the Carbon action coordinator, Mark Dowell.

# Agency and Liaison Reports

## US Geological Survey (USGS)

Kristi Kline gave a report on USGS activities at the EROS center. The USGS defined three basic categories of products: NRT products that are processed using ancillary data such as predicted ephemeris or bumper mode parameters that may be improved by reprocessing; Tier 1 products that meet the criteria for the collection definition; and Tier 2 products that do not meet the criteria for the collection definition and have been processed using the best known ancillary data.

Kristi summarized the findings of their Collection Definition Study saying that radiometric variability is not a factor and geodetic accuracy varies by sensor, source data type, the quality of PCD, and level to which the data have been processed (L1T, L1GT, L1G). She also described the Land Change Monitoring, Assessment, and Projection (LCMAP), which is a capability to continuously track and characterize changes in land cover, use, and condition and translate such information into assessments of current and historical processes of change that can serve as the science foundation that supports evaluations and decisions relevant to resource management and policy. USGS aims to modernize access to the Landsat archive, set the foundation for a federal land monitoring system, continue long-standing USGS land cover commitment, meet USGS land change science needs in terms of land change, geographic research, and the combined impact of climate and land use change.

For USGS, Analysis Ready Data means the data are processed to a level that enables direct use in applications:

* Allows geospatial, multi-spectral, and multi-temporal manipulations for the purposes of data reduction, analysis, and interpretation.
* Consistent radiometric processing scaled to TOA and surface reflectance.
* Consistent geometry including spatial coverage and cartographic projection – e.g., pixels align through time.
* Metadata of sufficient detail on data provenance, geographic extent, scaling coefficients, and data type.

Kristi gave an example of how the analysis is done today; this method is cutting out a lot of the time the scientists spend prior to the actual analysis.

Kristi reported that USGS is still pulling Sentinel-2 data, though network access and speeds continue to be an issue.

Kristi also described the USGS Global Visualization Viewer (GloVis), which is a quick and easy online search and order tool for selected satellite and aerial data, developed in 2001. The GloVis has been a popular visualization viewer for searching massive quantities of imagery stored in the Earth Resources Observation and Science Center (EROS) data holdings. Through a graphic map display, the user can select any area of interest and immediately view all available browse images within the USGS inventory for the specified location and download or order on-demand products through the interface. The current GloVis redesign uses modern languages and image processing tools and enables GloVis to improve interface for downloading and for requesting on-demand products, work seamlessly with current EROS systems for retrieving data and for user downloads, and improve code architecture to accommodate future imagery evaluation concept. GloVis will be Open Source.

Kristi concluded with the Landsat mission timeline, adding that Landsat 9 remains on track for December 2020 launch.

Nitant asked why the atmospheric correction is not done on-the-fly. Kristi replied that it is for the US, but it is using a projection that will not work globally; the plan is to expand to global.

## National Aeronautics and Space Administration (NASA)

Dawn Lowe gave a presentation on NASA’s Earth Observing System Data and Information System (EOSDIS). She listed their data sources categorized by satellite, airborne, and in-situ missions, and application and Earth science research support.

Dawn described their Distributed Active Archive Centers (DAACs), which ingest the Level 0 data, and process and distribute it, among other functions, ensuring safe stewardship of NASA’s data.

EOSDIS elements and capabilities include:

* EarthData: the EOSDIS’ website (*Earthdata.nasa.gov*) provides a focal point for cross-DAAC and EOSDIS content and news sharing, and access to Earth science data and services.
* Common Metadata Repository (CMR): metadata catalog of NASA's EOS data and related data services (e.g. reformatting, pattern recognition) as well as collection-level metadata for ~ 25,000 data sets from broad international community. Essential underlying component for data discovery and access.
* EarthData Search Client (EDSC): easy-to-use access to EOSDIS services for Earth science data discovery, filtering, visualization, and allowing users to download and access Earth observation data
* Near Real-Time Capabilities: Provided by “LANCE” (Land Atmosphere Near real-time Capability for EOS) which produces products within three hours of observation. NRT capabilities are co-located with the standard science production facilities.
* Global Image Browse System/WorldView: GIBS provides access to full resolution imagery derived from NASA products; Worldview client allows users to explore GIBS imagery in a Google maps-like manner.
* ESDIS Metrics System (EMS): collects and reports on data ingest, production, archive, and distribution across all EOSDIS data centers.
* Earthdata Log-in (User Registration System): provides a centralized and simplified mechanism for user registration and account management for all EOSDIS system components.

Dawn showed diagrams of data archive volume size and growth, and quantity of products delivered over time and by country. Products for EOS continuity were listed, categorized by land, ocean, ozone, atmospheric, sounder, and CERES (Cloud’s and Earth’s Radiant Energy System).

Dawn described the EOSDIS cloud prototypes, with goals to capture performance and usage, work with providers, and develop software to utilize capabilities of commercial clouds. Hosting applications in the clouds offers the clearest path to improve efficiency.

Kristi asked what criteria was used to choose Amazon as the vendor; Dawn replied that they chose Amazon because NASA already had a relationship with them. Lizhe asked if there is duplication between NASA and USGS. Dawn said that there is, for now. Lizhe also asked about progress on the ECHO. Dawn said all the ECHO functionality has been moved to the CMR (Common Metadata Repository). It will be open source.

Ken asked about user restrictions and tracking; Andy replied that they only track name, primary discipline area, country, and email, and it is not tracked individually, but by country and discipline. He wondered how many agencies are requiring authentication; it may be useful to work toward an OpenID method. Quite a few agencies were interested. Kristi noted that USGS is working on OpenID; user must register, but then can use Facebook or other password. Registration itself could be cross-platform as well.

**Action WGISS-42-20:** Technology Exploration team to investigate user authentication interoperability.

## Japan Aerospace Exploration Agency (JAXA)

Masumi Matsunaga gave a presentation for the Japan Aerospace Exploration Agency. She began with the JAXA launch schedule for 2016-17 and listed their ongoing and future missions. Masumi also discussed the G-Portal, from which users can download the various satellite’s data free of charge (open and free, except for ALOS and ALOS-2). She highlighted the JERS-1 SAR Level 0 data and products, which can now be downloaded, including SAR Level 2.1, OPS VNIR Level 2, and OPS SWIR Level 2. The tool which can convert CEOS format to GeoTIFF or KMZ format is under development and will be released by the end of March 2017.

Masumi noted several major improvements to the G-Portal in the areas of web search, catalog search, and direct download. The web design changed to “one page” design from “step by step” design. Users can also search satellite data from other agencies and G-Portal-R can cross-search catalogs. Users can use FTP for direct download.

Masumi gave the tentative 2017 schedule for JAXA’s data distribution service, which would include a unified user authentication using LDAP to aid interoperability.

## Indian Space Research Organization (ISRO)

Nitant Dube gave a report on the Indian Earth Observation Programme. He announced the successful launch of the GSLV-F05, PLSV-C34, PLSV-C33 and PLSV-C32 satellites, and listed the currently operational EO missions for land/water, high resolution, ocean, and weather/climate. He also described nine EO missions, with their specifications and applications, which are planned for launch over the next three years.

Nitant described the Meteorological and Oceanographic Data and Information Services (MOSDAC) multi mission data repository for multiple applications. It includes calibration/validation data, in-situ data, forecast, tools and utilities, research and different levels of training. Trainees are set to solve 300+ problems in a number of applications: cyclone, sea state, cloud burst, intense rainfall, and weather forecast.

He also described VEDAS for land applications, which is in a position to reach more and more users, and is developing new types of applications. They are working closely with government departments to develop customized applications for them.

Andy highlighted the point about calibration/validation data and in-situ data; it is important to make these more usable and discoverable to the users.

## Global Spatial Data Infrastructure (GSDI) Association

Gábor Remetey-Fülöpp gave the liaison’s report on the last six months’ activities of the GSDI Association. GSDI was established in 2004 for academic, government, industry entities and individual professionals in geomatics. GSDI today has 38 Organizational Members from 20 countries and over 400 individual members from 55 countries with a high concentration in developing nations. GSDI has Special Consultative status with UN ECOSOC and supports the UN Global Geospatial Information Management (UNGGIM) initiative. GSDI promotes the Open Data Principles of GEO/GEOSS. GSDI is involved in SDI capacity building activities in several ways, and has liaisons and Memorandums of Understanding with several global organizations.

Gábor described the GSDI 15 World Conference; its themes are spatial enablement in the Smart Homeland, Smart Disaster Prevention, Smart Transportation, and Smart City; he listed the theme groups and keynote speakers.

Gábor reminded that at WGISS-41 he took an action to share information on WGISS activities with the GSDI community. A WGISS contributing paper was submitted. The WGISS paper was accepted for oral presentation in the most relevant session at the GSDI 15. GSDI’s liaison invited to the GSDI Council and Board Meetings to share recommendations related to potentials of closer cooperation with WGISS/GEO. The basic version of the presentation might be updated on behalf WGISS (based on inputs from WGISS-42 and GEO XIII Plenary).

GSDI recent activities and topics include the GEOSS Work Program 2017-19 Workshop, the GEOSS Common infrastructure (GCI) Workshop, and GSDI at the 6th Digital Earth Summit “Digital Earth in the Era of Big Data”. Gábor also highlighted upcoming events.

Gábor discussed [NASA WorldWind, the open source, 4D virtual globe technology](http://eurochallenge.como.polimi.it/), and Europa Challenge. He listed the top five projects: Earthquake signal precursors, Quake hunter – Earthquake activity visualization, MultiVIS Analysis Suite, SpaceBirds, and NASA World Weather. He described World Weather in detail.

Gábor listed the following conclusions:

* Local, national, regional SDIs and associated services are enabling tools for EO applications.
* GSDI is capable not only serving and supporting EO applications, but also capacity building, providing awareness raising and generating user feedback with special emphasis on user requirements.
* GSDI is continuing support the GEOSS data sharing policy, promotes the Open Data Principles.
* The theme ‘Smart Homeland’ of the upcoming GSDI World Conference in Taipei is matching many of the GA’s focused areas introduced at the WGISS-41 Plenary.
* The synergy provided by the combined use of EO and SDI technologies and services is inevitable for achieving the UN sustainable development goals.

Andy commented that CEOS was very happy to hear that Gábor presented the WGISS paper at GSDI 15.

## National Oceanic and Atmospheric Administration (NOAA)

Martin Yapur focused the NOAA agency report on the National Environmental Satellite, Data, and Information Service (NESDIS), and their strategic plan for ensuring reliability, richness and robustness of services with commitments, capabilities, and communities. He listed their partners in the global observing system, and discussed the upcoming launch of GOES-R. Geostationary satellites like GOES-R, observe Earth from an equatorial view approximately 22,300 miles high, allowing them to see from the coast of West Africa, to Guam, and everything in between. GOES-R is considered a game-changer for weather forecasting and for NOAA. It will monitor the earth in near real time five times faster and at four times the resolution of the current satellites. GOES-R is scheduled for launch in November.

NESDIS’ architecture of the future will develop a space-based observing enterprise that is flexible, responsive to evolving technologies, and economically sustainable. The philosophy of the new director is to try to determine the value of investments, and to work toward building, blending, and buying data.

## Geoscience Australia (GA)

Simon Oliver gave an overview of GA activities, beginning with a discussion of Australia’s Regional Copernicus Data Access/Analysis Hub. The hub is pulling S1/2/3 and rearranging it into 5 degree by 5 degree directories every hour, working to incorporate S1 precise ephemeris and S3-Marine from EUMETCAST into the process and working on ARD S2.

Simon discussed the Australian Geoscience Data Cube (AGDC) v2 Programme, giving an update and outlining activities in the pipeline. He described the foundation concepts of properties of data, and datasets. He also listed the product’s supported and tested projections currently in the AGDC. Simon outlined the generalized analysis flow, adding that sensor ignorance is under construction; sensor ignorance is basically spectral equivalence.

GA is working with OGC on standardizing Discrete Global Grid Systems (DGGS), which is an important part of the future versions of data cube. The diversity, incongruity and lack of standardized applications of global grid infrastructures limits the development of accurate analysis tools for Big Earth Data. DGGS are a Digital Earth reference model designed to be information grid systems, not navigational grid systems.

Simon also reported on the GA Alice Springs Ground Station antenna upgrade and other recent projects at the ground station.

Simon noted that WGISS involvement would be welcome for data cubes, Discrete Global Grid Systems, and spectral response repository/standardization.

Andy wondered what WGISS or CEOS agencies can do to help with developing best practices for data cubes. Simon responded that WGISS can help ensure the capture of all the content available and work on data provenance.

## National  Institute for Space Research (INPE)

Lubia Vinhas began her presentation with a reminder of INPE’s concept of fostering data for the public good, and also outlined the status of the CBERS Program (missions, cameras, processing levels, and internal accuracy). She summarized these, adding that the conclusions listed are based on the comparison of resulting RMSEs with commonly accepted cartographic standards.

* MUX L4 images are suitable for mapping at scales 1:50,000 and smaller. MUX L4 images are extremely consistent in time in terms of their geometric internal accuracies
* WFI L4 images are suitable for mapping at scales 1:250,000 and smaller. Although WFI L4 images have acceptable geometric internal accuracies, a refinement in the optical distortion model of the two optical systems of the camera is still being analyzed.
* PAN5 and PAN10 L4 images are suitable for mapping at scales 1:50,000 and smaller. PAN5 and PAN10 L4 images have acceptable geometric internal accuracies that are about to be improved by the application of optical distortion models provided recently by their Chinese partners.
* IRS L4 images are suitable for mapping at scales 1:100,000 and smaller. IRS L4 geometric internal accuracies are not as acceptable as they should be due to inaccurate modeling of its camera push broom system.

# Data Use

## Future Data Architectures

Andy Mitchell gave a status report on the CEOS Future Data Architectures activity. He began by mentioning opportunities and challenges around the operating environment. These include the ability of developing countries to realize the potential value of satellite EO for the big global agendas and the desire to have solid concrete opportunities to develop partnerships with development banks and UN institutions. Other challenges are supporting next generation of climate applications, and promoting uptake by industry/value-adders, working together to lower technical barriers to enable industry to really get to work - ideally in a way that supports the CEOS concept that users having access to an international constellation of systems.

The study outcomes that are envisioned are an inventory of relevant initiatives and plans being undertaken by CEOS and related agencies. A report on lessons from the early prototypes with the governments of Kenya and Colombia is underway, as is a report identifying key issues and opportunities resulting from the trend towards Big Data and Analysis Ready Data. Another outcome is a list of recommendations for the way forward for CEOS and its agencies, including in relation to standardization, interoperability, and how the current CEOS priorities might benefit from the proposed activities.

Andy listed an outline of the report that is being developed. It includes sections on current trends and developments in EO systems architecture and applications, the challenge and opportunity of changing user expectations, increasing EO data volume, variety and velocity on EO systems architecture, the Future of EO Data Architectures, and conclusions and recommendations.

Andy reported significant activity across CEOS agencies with great diversity of approaches and capacities; FDA effort is needed, especially with the move to online data systems plus increased size/complexity of data. The big data players and their advanced platforms, populated with CEOS agency data (amongst others), are changing expectations as to how easy it could and should be to access and apply EO satellite data. Expectations are also changing due to the broadening user base for EO satellite data to more sectors and more users, many non-expert and not from large technical institutions; these have implications for ease of data handling for CEOS agency mission data. CEOS is placing more emphasis on supporting uptake and application of data – including grand themes like SDGs, climate and food security.

The future concept is an interaction model for in-place analysis of ready-to-use data is replacing discovery and download. It dictates architecture changes in the interfaces between agencies, between computational infrastructures and agencies, between discovery and analysis tools and users. There are significant and diverse approaches among agencies, including bringing the user to the data, APIs/Virtual Laboratories. These are enabled by standards, by pre-processing data to a point where it is a measurement comparable in space and time with other measurements from other sectors, by integration of different types of data across different domains, moving the burden of data processing for the extraction of application information from the users to the space agencies (TEPs, ARD..), and by HPC approaches such as the CEOS data cube.

The current situation is a major strategic challenge for CEOS: Technology is changing rapidly, and agency initiatives do not happen in a vacuum. There will be no ‘single system’ or ‘single stack’ solution, so it is important to work together to take advantage of these opportunities to highlight the value of CEOS’s efforts in coordinating ‘virtual constellations’ that reduce risk and add simplicity for users, and help users benefit from all CEOS data.

The next steps are for the FDA ad-hoc team to break down recommendations as short, medium and long term, with short term recommendations in the 2016 Report to CEOS Plenary. USGS and CSIRO are willing to take forward the effort through 2017, subject to CEOS Plenary approval, focused on medium- and long-term recommendations. The team has a pilot, prototypes and enabling technology work underway within SEO, LSI-VC, GFOI, and WGISS and the team recommend these be progressed and accelerated.

WGISS efforts on Discovery Search Engine Optimization, Access Common Standards for Interoperability of Product Formats (metadata/data) and API, and exploration of emerging Big Data services including Cloud Computing should continue.

CEOS FDA-related efforts include:

1. Small-scale demonstration of the potential of ARD
2. Continued development of CEOS data cube as an example
3. Explore other FDA-related suggestions
4. Extend the FDA ad-hoc team at Plenary for one year under USGS/CSIRO leadership

Andy also discussed Analysis Ready Data, which is separate from FDA study but feeds into it. The LSI-VC were tasked (Nov 2015) to “Define intercomparable Analysis-Ready Data (ARD) products within the context of land surface imaging”. The CEOS Analysis Ready Data for Land (CARD4L) is defined at two levels, a general description and technical specifications:

Minimum requirements for CARD4L are:

* General metadata lineage; authoritative data
* Quality metadata quality flags; suitability for a particular use
* Corrections allow the users to use the data directly; geophysical measurement of the land surface
* Geometric correction accurate location; data can be ‘stacked as time-series’
* Optical: atmospheric and solar view angle corrections
* Radar: radiometric corrections for topography and incidence angle

Andy described the CARD4L specifications and benefits. The challenges are the closure on the high level definition/description document is needed to allow LSI-VC to focus on the next level of detail; increased participation is sought from agencies to lead and participate in sub-teams to take this work forward; specifically, a team is needed to work on specifications for radar data, with involvement of the major radar providers.

A discussion of the draft report followed. Kristi remarked that the report seems to focus on access and analysis, bringing users to the data; it touches on topics and then does not follow through. Richard commented that the user needs and classification of users is missing from the document, and there is not much about distribution and catalogues. Clarification is needed on the topics of cost for data and services, and cost of cloud computing. Standards and recommended methods need to be mentioned.

Andy asked if the study report is a document that participants would find useful in their agency. Lubia responded that they do recognize that it is important, but do not see its applicability; the general ideas are there, but the specifics are missing. It is too broad in some areas and too specific in others.

Satoko commented that the target is very unsure. When JAXA received the document it was handed down to people for whom it was too general. The target audience needs to be clarified and solutions need to be clarified and detailed. Satoko suggested that the report from the Cloud Computing Workshop may serve as an appendix for the FDA report.

Andy listed a few other comments: it is focused on land, it is not clear if the scope applies to in-situ users, and the appendix needs to be cleaned up.

Brian agreed that it is important that WGISS have a significant input into this report; the report is a large undertaking.

**Action WGISS-42-19:** Technology Exploration team to create a summary report of the Cloud Computing Workshop; the summary should be tailored for input to the FDA report.

**Action WGISS-42-18:** WGISS-42 participants to review Satoko Miura’s outline of the Cloud Computing Workshop and provide feedback.

## Data Cube Projects at ESA

Guenther Landgraf gave a presentation on data cube projects at ESA. He described a new way of working, where the

* traditional user operates as: Access🡪 Download🡪 Exploring🡪 Processing🡪 Results
* EO data cube user operates as: Access🡪 Exploring🡪 Processing🡪 Results🡪 Download.

Guenther displayed a diagram of the European optical HR data cube infrastructure.

The Open Geospatial Consortium (OGC) [Web Coverage Service](http://en.wikipedia.org/wiki/Web_Coverage_Service) (WCS) interface standard defines web-based retrieval of coverages – that is, digital geospatial information representing space/time-varying phenomena. He displayed examles of WCS and WCPS.

ESA’s user uptake promotion plan includes two pilot projects: Land Productivity Map Production (ENEA Italy), and Urban Growth Monitoring in Eastern Austria. Several questions are in discussion: Are level 1 data suitable for new access technologies? Is the user comfortable with WCS/WCPS? Extend to python?

The EO exploitation platform data cube includes storage/access of “features” (analysis-ready-data”) produced by Thematic Exploitation Platforms. The Central European Facility is shared by all exploitation platform activities. Procurement is expected to end in 2017.

Philippe Mathieu spoke next on Earth System Data Cube, where each community wants to connect their data, so each community puts their data on the same grid. This grid has global extent, nested spatial grids, convenience aggregations, and consistent temporal sampling; it gives priority to the ESA data suite, and includes uncertainty information. He listed the essential variables from atmosphere, biosphere, and anthroposphere. The effort is towards biosphere-atmosphere-society system trajectories, and building a narrative on how the Earth system is moving.

Kristi noted that agencies start with very high level products, but hope to converge to lower level ones. In response to questions from Nitant and Lizhe, Philippe noted that this project has been a quick exercise up till now, so more maturity is needed. For now, the user can download subsets of the data in the cube. The approach is a big data analytics approach; the cube will always be updated with new data as it comes in. The user can just use their own python framework.

## Open Source Big Earth Observation Data Analytics at INPE

Lubia Vinhas presented an INPE project in the direction of a data cube: an open source big EO data analytics project for distinguishing forests by temporal evolution. Two paradigms are considered: space first, where images are classified separately, and results are compared in time. Or time first, where time series are classified separately and results joined to get maps.

The desired analysis is the transition from natural to managed land as seen by remote sensing, considering seasonal, double cropping, single cropping changes, and forest degradation and deforestation events. Time series mining depends on pattern-matching: matching land use patterns in a remote sensing time series. One approach is using events to understand land transitions, and interval temporal logic.

Lubia described and illustrated land use change trajectories in the Amazonian biome of Mato Grosso using MODIS x time-weighted DTW, using middle-resolution time series (landsat-8 plus Sentinel-2A).

INPE applied for and got a grant for a project to acquire new insights into land use change, and to use new data analysis methods. The project will use an array database for big scientific data and free satellite imagery. It allows scientists to check the algorithms, proof of concept prior to processing for entire data set. The project is not at the INPE organizational level, but an attempt to show the need for something like this for data cubes.

In response to a question from Guenther, Lubia said that the projects is being carried out mostly by students, but they have exposed this to the scientists. The first year there was a lot of work to do to put the project in place; now it is time to expose it to more senior researchers. It is hoped it will be an environment that the scientists will want.

Brian commented that these last two presentations show examples of data cubes, and the interest in time-series. Over time there will be multiple ways to get to the solution. Flexibility is key, and maybe some common lessons learned and common algorithms, where WGISS takes the lead.

Martin asked Lubia if she sees INPE formalizing the funding for this. Lubia replied that the method of obtaining these products is very appealing to the ministries, but budgets are tight, and INPE still has to continue their normal work.

Discussion on how to make these initiatives more interoperable is needed.

## Use Discussion

Gábor Remetey recommended the introduction of GSDI operational applications using EO data, to demonstrate applicability to the geospatial community. He offered to raise this at the next GSDI board meeting, and added that a geospatial applications workshop that incorporates the users would be a good outreach for WGISS. A “client” day workshop in which the users demonstrate their clients could also provide user feedback.

**Action WGISS-42-14:** Gábor Remetey to coordinate (via WGISS) GSDI contributors to a future WGISS geospatial applications workshop on the use of CEOS data.

# Data Access

## International Directory Network (IDN)

Michael Morahan reported on the IDN. He began with the recent updates to the IDN home page and to the keyword search interface. The objectives of the update to the search interface are a one page search, refinement, improved search precision and recall via new functionality design, interdisciplinary search capabilities through the selection of multiple facets, and enhanced sorting capabilities by allowing users to add sortable columns to the search results area.

The GCMD/IDN CSW service will be replaced by a new CMR CSW service this year. A forwarding service will be set-up temporarily to send users to the new service.

Michael discussed the GCMD keywords describing the requirements for additions/changes to the IDN hierarchical set of controlled keywords covering the Earth science disciplines. He listed the keyword types by discipline, and how they are structured (by discipline, platform, data center, and location). He described the governance and implementation processes in detail. The keyword Community Forum page shows information for these.

IDN metadata are assessed to ensure that they are compliant, accurate, complete and intelligible (CACI) for effective data discovery and access. The metadata support development of standards, rules, and tools to enhance curation effectiveness and efficiency. Michael described the process for performing quality assurance (QA) using automated rule checks and manual review, and the process for making recommendations to providers. Both automated and manual metadata checks are needed to ensure high-quality metadata. Automated checks can free up time to focus on making fixes and to identify where additional manual review is needed. When the automated QA process first started, it took over a week to generate a report; now it takes hours. QA Triage reports can inform providers of recommended changes.

Michael reported IDN usage metrics and metadata record counts.

Nitant requested the advantages for migrating from DIF-9 to DIF-10. Michael replied that it is not required, but is more in line with the new data model, and provides more fields, has more structure, and many of the fields are ISO Compliant. A tool will do the conversion, but not all the metadata is in DIF-9, so some manual updates will be needed. DOIs are not a requirement, but can be easily added to the DIF-9 or to the DIF-10.

Mirko asked if DIF-10 will be stable. Michael thought it would be quite stable at this point. There will be no more updates to DIF-9, but DIF-10 will be maintained and updated.

In response to a question from Uwe Voges, Andy said that the DIF-10 is being mapped to the UMM and the UMM does have an API. The DIF-10 has better search relevancy and the MMT (metadata management tool) allows the user to output the ISO. The IDN accepts ISO DIF CMR metadata.

Simon asked about search functionality; can the user identify a search of one sensor similar to another. Andy replied that the Federation of Earth Science Working Group has discussed this is a topic.

Richard asked if the IDN references services; Michael replied that the team is in the midst of a new model for services. Richard noted that the webpage refers to services, and suggested removing it. Richard added that it is interesting to put service endpoints and maybe it is more efficient to have it in the IDN. Mirko asked that the UMS be circulated, and Andy agreed to do so.

Martin asked if NASA support for the IDN is stable. Andy replied that it is; since NASA unified the CMR to the IDN, NASA is highly dependent on it. It is a requirement for NASA to have science keywords in all their metadata.

Andy added that CEOS has an open action for CEOS agencies to ensure that all their metadata is in the IDN; GEO is querying the IDN for CEOS records.

The team encouraged the participants to read the keyword governance process to provide feedback.

**Action WGISS-42-12:** WGISS to explore to possibility of including publishing services related to EO datasets in the IDN.

## CEOS OpenSearch II Project

Yves Coene gave the background on the OpenSearch Project, and described the agreement to separate the Developer Guide (DG) from the Best Practice (BP) document. He explained the version evolution of the Best Practice document and listed the requirements that were moved from the Developer Guide to the Best Practice document. The CEOS Open Search Best Practice was released prior to WGISS-42. The CEOS OpenSearch Developer Guide version 2.0D2 was released before WGISS-42, and is available for review. The CEOS OpenSearch Conformance Tests are to be discussed.

Yves reported that the OGC has published OGC 10-157r4 on the public website; it has no impact on the CEOS OpenSearch BP/DG documents. OGC TC and PC have approved OGC 13-026r8 (EO extension for OpenSearch) dated 06 July 2016; the CEOS BP and DG are based on OGC 13-026r5.

The impact of OGC 13-026r8 on CEOS OpenSearch BP/DG documents is:

* CEOS-BP-011 no longer needed as it is now mandatory in base specification.
* Changes to recommended media types defined in atom:link as per table 7 in BP v1.1.

Andy asked if the impact of the changes to the new OGC standard have been distributed to WGISS. Yves said he could easily distribute the document. Mirko asked if the CEOS Open Search Best Practice can be said to have been accepted and approved by WGISS. Since referencing the OGC is a minor editorial change, it can be considered as accepted and approved by WGISS. Mirko suggested one more month for comments on the Developer Guide, and then issue the acceptance at WGISS-43.

Andrea Della Vecchia suggested proposing a new activity for conformance testing. Yves said that NASA has a nice tool for this, and Andy confirmed that it will it be released as Open Source.

**Action WGISS-42-05:** Yves Coene to distribute to WGISS the new OGC Specification for OpenSearch.

**Action WGISS-42-06:** OpenSearch team to finalize and post the CEOS Open Search Best Practice after making minor editorial changes referencing the OGC.

**Action WGISS-42-07:** OpenSearch team to circulate the Open Search Developer’s Guide after allowing one additional month for comments.

## OpenSearch for EO Evolution

Olivier Barois gave a presentation on the evolution of OpenSearch for EO. He noted that the OpenSearch for EO Standard [OGC 13-026r8] was recently approved at OGC, and is about to be published. There is improved “interoperability” between EO catalogs, and it is much simpler and cost effective to implement than previous standards. It has been adopted by CEOS/WGISS (FedEO, CWIC) and by many EO providers. It is a success story, but is tailored for RSS clients (results in Atom format). The EO metadata model (the details of the search results) are not really “interoperable”, and not really optimized for web clients. There are many different alternatives of json/GeoJson base encoding.

ESA and EUMETSAT agreed to join their effort to work on a future evolution of the OpenSearch Extension for Earth Observation [OGC 13-026r8] standard. The main goal is the adoption at OGC of OpenSearch for EO GeoJson encoding standard, to define a vocabulary, propose an HTML encoding enriched with linked data information, and propose a json-ID context to wrap the GeoJson encoding.

Olivier noted that the preliminary study is collecting inputs. The presentation of standard candidates will be at OGC TC in Delft (NL) in Mar-2017.

Olivier reported a use case of distribution of search results and displayed catalog search response as a GeoJson OWS context. He displayed the mapping of product metadata attributes, and an example for Sentinel-1 metadata.

Jérôme commented that this is an excellent idea and CNES has been experimenting with this. It should achieve something really useful to present to OGC. Uwe suggested calling it HTML encoding, rather an annotation. There is a risk: it cannot be expected that Google will index the huge amounts of representations, but the reference can be properly indexed.

Olivier mentioned a general tendency to do json encoding so they are following the trend; it is a good opportunity to serve the community.

**Action WGISS-42-08**: Mirko Albani and Olivier Barois to research using the GEO-JSON encoding, and identify members that can participate in the OGC group.

## Federated Earth Observation (FedEO)

Yves Coene presented recent activities and achievements of FedEO. He reported that the eoPortal client is evolving. The main changes include restyling, making suitable for discovery, download and ordering, autocomplete based on OSDD info, metrics portlet, and download granule/collection metadata in various formats.

The service monitoring objectives are to provide online tools for the administrator to monitor service back-ends, and to include service information in dataset series search responses (Atom) to be exploited by the FedEO clients. He discussed the service monitoring console and response, adding that the OpenSearch access has been added to metrics information.

Other ongoing work is to include the evolving standards, to allow obtaining metadata from gateway and client, and to enable alternative and Linked Data response formats as being deployed on OBEOS Gateway server.

Catalog and data access updates include metrics that can be generated automatically to a spreadsheet, three new collections (OceantSat-2, SMOS Open, and CNES PEPS) have been added, and SSARA has been reorganized for better searchability. The collections of ESA CDS (including Proba-V, ROSCOSMOS) have also been increased.

Conclusions:

* Collections accessible through FedEO (8/2016)
* Metrics evolution: Harmonization Collection Metadata format (OGC 11-035r1) and Product Metadata format (OGC 10-157r4)
* Two-step search: local FedEO Collection Catalog is now OpenSearch (SolR-based) replacing I15 EP.
* Collections per endpoint were recently added or updated.

Ken noted that there is probably overlap with the NASA/CWIC metrics; Yves said they are working on cleaning that up. Richard asked if the intention for ESA is to use json and Linked Data. Yves replied that they have applied the standards, but the mappings are not standards, and may be inputs to a standard study. Olivier added that this effort is not a strategy; right now they are experimenting. Andy commented that he really liked the design of the new eoPortal, and it serves as a good model.

## CEOS WGISS Integrated Catalog (CWIC)

### CWIC Report

Yonsook Enloe reported on the CEOS WGISS Integrated Catalog. She explained that the IDN has been migrated to the CMR, and this is synchronized with CWIC. She described the four-stage publishing of new NASA datasets (registration in IDN/CMR, synchronization in CWIC, testing in CWIC, and tagging ready-to-search). IDN serves as a first step in the two-step search. CWIC accessible is tagged in the CMR as such.

Yonsook listed the registration status of the various NASA datasets; there are 4242 datasets ready-to-search (about 2500+ new datasets have been added via IDN/CMR and about 700+ datasets to be added to the operational service soon).

The CWIC connected assets total 4451 datasets and about 258,765,892 granules. Datasets to be added in the future include: NOAA NCEI (many thousands of datasets/ many millions of granules), ISRO (new missions), China (3+ data centers), Australia (NCI and BoM), Russia (ROSCOSMOS), and more.

Yonsook displayed statistics of successful data requests metrics in the last two months. Use of the CMR Concept ID is growing, but they will continue to support the GCMD Entry ID as long as it is needed. Yonsook reported that the CSW is now supported by IDN/CMR for collection-level queries, and the CWIC/ IDN (CMR) synchronization is fully functional. The CMR concept-id is now persistent (safe to cache) and some GCMD EntryIDs have changed, which could cause problems for GEO CSW clients that have cached old identifiers.

Yonsook described a guide to update the dataset id, and noted that the Dataset Identifier changes will result in correct identifiers for dataset queries with CSW Capabilities. The team recommends the use of CMR concept-ids when at all possible, and to update cached identifiers or DIFs.

Yonsook reported that the CWIC data partners are revising their internal systems: USGS and INPE are implementing OpenSearch, NOAA is re-implementing their system for GHRSST and then later for NCEI data, and EUMETSAT is offering access to their operational database. In addition, AOE, ROSCOSMOS, and Australia (CSIRO, GA, BoM) are working on 1.0.

Yonsook concluded that all connectors are now compatible with both CMR concept-id and GCMD EntryID as identifiers. The USGS/LSI connector supports USGS CEOS OpenSearch implementation, the EUMETSAT connector is now using the EUMETSAT production server, and the INPE connector is now supporting CBERS-4 datasets (can manage multiple server URLs for single data center).

Michael noted that the EntryID must be unique.

### EUMETSAT Report

Uwe Voges gave a presentation on the EUMETSAT CWIC status. He began with a list of new EUMETSAT collections provided to IDN / CWIC. The collections were tested and documented, and provided with DIF metadata for IDN. They also provided Sentinel-3 dataset metadata and let it register with CEOS IDN.

Uwe described the EUMETSAT OpenSearch (EOPOS) Interface. They updated the following links in EOP search interface (EOPOS) responses: USC (EUMETSAT Order) client, and EO Download Service (OGC 13-043). He added that random tests to check consistency between temporal coverage of DIF metadata with provided products were conducted, ignoring large spatial filter in EOPOS. EOPOS is now enabled on EUMETSAT OPE server. Uwe described some issues that they are working to resolve with CWIC.

Uwe explained that they are also testing with the CWIC Smart Client, and listed the EUMETSAT OpenSearch (EOPOS) interfaces. He raised the following questions:

* What are the plans for getting an operational site; how is the workflow Dataset search from IDN – Granule search in CWIC
* Have you got any monitoring/reporting statistics of the usage of the current system
* Do you want to support OGC OpenSearch-EOP 1.0 and the planned OGC OpenSearch-EOP 1.1 with GeoJson binding? Would NOAA/NASA want to be part of the SWG and support this?

Yonsook described how the two-step search is done, through CSW and through OpenSearch. Now that CWIC has a significant number of partners, they are in the process of describing this for those working on clients for their search and access. She also noted that the NASA data search tool will be the CWIC client and there are plans to make it Open Source.

### ISRO Report

Nitant Dube gave a presentation on the ISRO CWIC status. He began with an overview diagram of the ISRO connector interface with CWIC, and reported that the MOSDAC and NRSC data centers are interfaced with CWIC.

The Meteorological and Oceanographic Satellite Data Archival Centre (MOSDAC) is operational using a CSW interface; data is free, only registration is required. INSAT-3D IMAGER and Kalpana-1 VHRR metadata and products are available.

The National Remote Sensing Centre (NRSC)is operational using an OpenSearch Interface. Data from Bhuvan NRSC Open Earth Observation Data Archive (NOEDA) products, IMS-1 Hyper-spectral data, and Oceansat-2 GAC L1B products are available through CWIC via direct download, and are free. Commercial data (Resourcesat-2 LISS-3, LISS-4 and AWiFS data) are on chargeable basis and require registration.

Future plans include the registration of MOSDAC INSAT-3DR and CARTOSAT-1 DIFs and of more DIFs from NRSC, and to explore the possibility of enabling of OpenSearch for the MOSDAC connector. Finally, ISRO is working on OpenID-based authentication, beginning with users coming through the CWIC link, to enable and encourage CEOS users to use CWIC.

### INPE Report

Lubia Vinhas gave a presentation on the INPE CWIC status. She began with a CBERS program status, noting that CBERS-4 is fully operational and images are available for download. She explained the CWIC architecture and gave the links for their inventory service. She reported that they can access the CWIC connector, and have implemented OpenSearch.

Next steps are to complete the testing and integration of CBERS-4 dataset on the CWIC Connector, to complete the OpenSearch endpoint considering the EO Standard and the Best Practices documents, and to finish the registering the CBERS-4 datasets:

* INPE\_CBERS4\_AWFI
* INPE\_CBERS4\_PAN5M
* INPE\_CBERS4\_PAN10M
* INPE\_CBERS4\_IRS
* INPE\_CBERS4\_MUX

Lubia requested official feedback of the visibility of the data; this would include metrics showing that CWIC is providing visibility outside of Brazil and will help to justify the support of this activity. Andy noted that he has heard this from other data centers; the difficulty is that it is hard to do with brokers, but suggested a discussion on how to discover the lineage. Nitant reported doing this based on data and products ordered from CWIC, but this does not provide search metrics; Andy added that the problem will increase with the cloud. Ken agreed that it is more than just metrics; the ESIP world is encouraging citation of the data source and provenance.

### NOAA Report

Ken McDonald gave a presentation on the NOAA CWCI status. The NOAA data provider for CWIC is the Group on High Resolution Sea Surface Temperature (GHRSST). NOAA is continuing to investigate access to the full set of NOAA satellite data resources held in the Comprehensive Large-Array Stewardship System (CLASS).

Ken reported that along with NASA, NOAA is supporting CWIC development and operations. A task is in place with a team led by Dr. Liping Di through end of 2017. CWIC provides infrastructure services for brokered discovery/search/access to diverse set of satellite data from a set of data provider partners. Both CSW and OpenSearch clients are supported, and a rich set of data partners exists. CWIC is also supporting operations. CWIC clients are:

* Generic CWIC Clients that provide common discovery/access to all CWIC partner holdings, and fixed set of search/access services across all collections.
* Specialized Community Clients/Portals that limit search domain to specific discipline or area of interest, enable integration of CWIC satellite data with other data types, and provide other enhanced search/access services.
* Combination of Generic and Community Clients that maximize utilization and benefits of CWIC infrastructure

As part of its support to CWIC Dev/Ops task, NOAA added a subtask and resources to design/implement CWIC community portals. In this initiative, they will review existing client/portal implementations, identify one or more communities that wish to collaborate on the initiative, engage community representatives to establish initial set of requirements, design portal architecture in collaboration with community, and implement a prototype. So far, the team has reviewed existing portals (LSI CWIC Portal, NASA Earth Data Portal, and CEOS Water Portal), and also GEO Community Portal activity.

The CWIC client initiative is similar to the GEOSS Community Portals Subtask under the GCI Development (GD-07) Task. The GEOSS CP Team is comprised of GEOSS Common Infrastructure (GCI) Providers and Community Representatives. The idea is to maximize utilization of GCI by promoting and enabling development of community portals/clients and leveraging community infrastructure capabilities.

While initiated within the CWIC Project, portal implementation should investigate access to CWIC, FedEO, GCI and other providers. Enabled by CEOS OpenSearch Protocol, these can provide common access to all CEOS holdings, and GCI opens up broader access to in-situ data. The team proposes to extend services beyond discovery and access to visualization and animation services, simple analytic functions, on-demand preprocessing and customization (e.g. re-projection, re-formatting, etc.), and service requirements established through interactions with communities

Ken noted that Andy and Mirko proposed a portal collaboration with the WGClimate (carbon) at the SIT Workshop last week; it had a positive initial response. Follow-up with technical representatives is needed/anticipated.

Yonsook commented that the community portal work is perfect timing. The data partner infrastructure is set up, and several search clients can be adapted to a number of communities. Andy noted that there are other clients that were not mentioned; WGISS should track those clients. Ken added that they are also looking for partners.

Andy reported that he had a discussion with the VCs as to whether they need a thematic portal, and is awaiting feedback. Another idea was to have a workshop helped by WGCapD to see what the developers’ needs are for developing a client. Kristi recommended a client where one of the first steps is to select the data and store the searches so that every time you log in it brings up previous search criteria.

Andy noted two things from the SIT meeting:

* Need to work with WGClimate (carbon portal)
* Work with GEO Portal for climate

Ken said he would be happy to pursue those two activities, and these should be added to the work plan for the Carbon portal.

## WGISS Connected Data Assets

Andrea Della Vecchia gave a presentation on the WGISS connected data assets effort. He began with a description of CWIC, FedEO, the IDN, and WGISS interoperability standards, and also displayed the WGISS Connected Data Assets web page.

Next he discussed the connected data assets architecture, noting that it is difficult to understand outside of WGISS. The following issues have been identified, which lead to the need for a unified architecture for WGISS assets:

* Duplication: E2E architecture risks to return the same result twice (e.g. GEO DAB may route a request both to CWIC and FEDEO and get NASA CMR or EUMETSAT results twice). User risks getting many collections that are included in other collections
* User might get different results when accessing different portals
* User has no idea about stability (e.g. DOI) and quality of collections
* Ranking of results might not match user query and does not take into account where the user might actually have higher rights to access data
* Infrastructure leaves user “alone” for access to the real data (after discovery); discovery and access to data associated knowledge and information is not linked to the data
* Different metadata models and transport layers do not make the user’s life easier… and implementation costly

A WGISS Integrated System team needs to be set-up to coordinate and oversee the WGISS Integrated System and Standards. This team would:

* Coordinate operations, maintenance and evolution activities (e.g. for infrastructure, standards adoption, etc.)
* On-board new data partners
* Provide technical support for client partners
* Monitor the health of the federated system and report outages and errors to the partners
* Test all the components of the federated system, including end-to-end search and data access
* Work with data and client partners to identify and resolve system and component bugs
* Provide support for metrics collection

Future activities:

* Review CEOS discovery and access infrastructure architecture, functions and interfaces.
* Concerted approach to give homogeneous results to the user allowing to easily filter duplicates within the CEOS infrastructure
* Ensure reliability and redundancy of CEOS discovery and access Infrastructure (no single points of failure)
* Address shared collections and mirror collections
* Harmonize collection/product metadata models along ISO 19115
* Enhance directory information with data access condition information and quality indicators
* Provide result ranking with user-selectable criteria
* Link data to associated information and knowledge for user access
* Harmonize terms and conditions acceptance and data access authorization process in a federated user management approach
* Improve minor shortcomings on current standards (e.g. OpenSearch, json binding, …)

Andrea proposed a two-phase approach:

Phase 1: consolidation of current CWIC/FedEO/IDN overall architecture to quickly address some of the identified open issues (Q3 2016 – Q2 2017):

* Consolidate overall architecture and concept (starting from previous slides)
* Consolidate components and interfaces to partners
* Remove duplicated discovery and access to identified data collections (e.g. EUMETSAT, NASA, ASF, ROSCOSMOS) to align with overall concept
* Consolidate and maintain WGISS web pages and a coherent set of statistics
* Finalize OpenSearch Developer Guide and Best Practice
* Contributors: CWIC/FedEO/IDN teams, WGISS support

Phase 2: Dedicated “Interoperability Project” to evolve and fully address all open issues (preparatory phase Q3 2017, implementation phase Q4 2017 – Q4 2018).

Yonsook agreed with these suggestions, and Ken complimented Andrea on his presentation, saying that all these points are key; the whole idea of brokers really needs to be addressed.

Katrin asked about the duplication caused by GEO and CEOS results; Andrea responded that first step is to find a solution to this problem within the CEOS architecture. Andy noted that WGISS has discussed using DOIs as one way to solve it. Mirko suggested advising GEO to pull satellite data from CEOS only, and Andy suggested starting the conversation on best practices for this duplication problem. This can go into the GCI system requirements in order to get widespread distribution.

**Action WGISS-42-09:** WGISS representatives to volunteer (by sending email to Andy Mitchell and Mirko Albani) for the WGISS Connected Data Assets System Level Team.

**Action WGISS-42-10**: Yonsook Enloe to initiate a teleconference for the first meeting of the WGISS Connected Assets System Level Team.

**Action WGISS-42-11**: Yonsook Enloe to revise the GCI User Requirements document to insert draft requirements addressing the problem of harvested/cached datasets and duplicate datasets in GEOSS.  Initiate a review of the new draft GCI User Requirements with members of WGISS.

## Unified Metadata Model: WIGOS/CGMS Mapping

Simon Cantrell gave a presentation on a Unified Metadata Model (UMM) with the purpose of providing a common metadata model to unify legacy systems (i.e. GCMD, ECHO) with new systems (i.e. CMR). The UMM supports interoperability of legacy systems, and additionally supports interoperability of other national and international systems. It contains metadata representing several key entities associated with EOS data: Collection, Granules, Variables, Services, Visualization. Simon displayed a diagram representing granule metadata in the UMM, related to the collection metadata, the variable metadata, service metadata, and future metadata concepts. The model specifies the basic characteristics of the observed variable and resulting datasets, and includes an element describing spatial representativeness of the observations as well as the biogeophysical compartment.

Simon compared the WIGOS UML Data Model and the CGMS UML Data Model with the UMM, giving a detailed analysis of mappings. Similarities between the models are with spatial, temporal, and acquisition information. However, variable level information in WIGOS is different, and product level information in CGMS is different.

Andy reported meeting with the WIGOS representative regarding the GEO Foundational Task GD-03 which involves WGISS. WIGOS was asking for input from the satellite community as they develop their standard; it would be nice to do a comparison and come up with a synthesis. He added that his team will be working on this, and may make a recommendation or perhaps a webinar where the community can discuss it.

**Action WGISS-42-13:** Andy Mitchell, Yonsook Enloe and Simon Cantrell to work with WIGOS and CGMS metadata representatives to identify common collection-level metadata elements between the IDN, CGMS, and WIGOS.

## Access Discussion

The participants were urged to think about discovering and accessing data in future architectures.

# Data Preservation

## Data Stewardship Interest Group Update

Mirko Albani gave an overview and update of the Data Stewardship Interest Group (DSIG). He began with feedback from the GEO Data Management Principles Task Force (DMP-TF). Mirko reported that the Data Management Principles Implementation Guidelines will be presented as an information document at the GEO-XII Plenary (Nov15). It contains a description/explanation of each principle, guidance on implementation, resources implications for implementing each principle, suggested metrics measuring adherence to each principle. Mirko and Richard are contributors, and harmonized feedback was provided by European Space Agencies in the frame of LTDP WG activities.

Mirko discussed GEO GD07.03 3 Year Work Program proposed subtask activities.

Mirko presented the CEOS Data Stewardship Best Practices Document Tree. Documents for adoption today include the Data Purge Alert Procedure. The procedure was endorsed at the CEOS Plenary 2015. The white paper was adopted by the WGISS participants. In case of Purge Alert notification, the WGISS Chair will inform all CEOS Agencies Principals and publish a ”Purge Alert News” on the CEOS and WGISS web pages.

Documents in progress include Associated Knowledge Preservation Best Practices, and CEOS Maturity Matrix. Mirko displayed their timeline.

Mirko discussed various cooperation activities with other CEOS working groups, virtual constellations, and ad-hoc groups. They have had some discussion with WGClimate dealing with the recovery of historical datasets, in response to a request to help in identifying additional datasets of interest for ECV generation which were acquired in the past but are not accessible today.

Mirko listed conferences of interest:

 2016 Conference on Big Data from Space, November 28-30, in Toulouse, France; a continuity of the previous year’s conference.

The Living Planet Symposium 2016 in May 2016; Summarized events and gave location of presentations and proceedings.

PV2017 will be November 2017.

Nitant suggested developing a video on historical data preservation, showing the recovery of datasets for WGClimate. Andy suggested putting this on the Faces of CEOS page.

**Action WGISS-42-15:** DSIG team to consider developing a video on the value of WGISS Data Preservation activities for the Faces of CEOS series.

## Preservation of Associated Knowledge Best Practices

Mirko Albani discussed the best practices of preservation of data associated knowledge. He began with the background obtained from previous discussions resulting in the recognition of the need for a more harmonized approach. This led to the drafting of CEOS Best Practice on recommended approaches for associated knowledge (i.e. information and software tools) preservation.

Mirko discussed the table of contents of the Best Practice that is under development. He described the comments received at WGISS-41, and the resulting clarifications that were made. Mirko reported that these have all been addressed.

He showed the recommendation ID scheme pattern, and discussed the recommendations for formatting and for software. Information preservation guidelines were listed, including information format recommendations. Software preservation recommendations were detailed by recommendations for future missions, historical missions, and current missions.

Andy asked if WGISS should be looking at making recommendations on open source software to make it a little more interoperable. Nitant did not think that there is much more to do on this. Mirko and Iolanda said they already have some guidelines outlined in the document, though maybe these should be taken one step further. Andy asked if this needs to be taken to the GEO Data Sharing group; NASA is working on an Open Source policy, and recommendations would be useful. The DSIG agreed to check with the GEO Data Sharing group to see if they have any recommendations for OSS standards.

The next steps for DSIG is to request new comments on the BP from WGISS members by the end of November. Comments implementation will be planned for February 2017, with subsequent circulation of final document (if mature).

Nitant asked if preserving the production guides is a part of this. Mirko replied that they have listed all the types of “knowledge”; that list does mention the log files, and the format in which they should be preserved.

## Maturity Matrix/Model for Harmonization

Iolanda Maggio discussed maturity models/matrices that are used to measure “levels of maturity” addressing the needs of specific domains. She gave the example of the scientific data stewardship maturity matrix (MM). The matrix leads to three questions:

* What is the maturity matrix/model: All activities needed to preserve and improve the information content, quality, accessibility, and usability of data and metadata.
* Who could use it: Data providers, modelers, scientists, decision-makers, data managers, stewards of data centers and repositories.
* Why should it be used: It provides data quality, usability information to users, stakeholders, and decision makers, and is a reference model for stewardship planning and resource allocation. The matrix creates a roadmap for scientific data stewardship improvement and provides detailed guideline and recommendations for preservation. It evaluates if the preservation follows best practices, while giving a technical evaluation of the level of preservation and helps with self-assessment of preservation with no numbers or average but a status. It helps to break the problem down, and understand the costs associated with each, so that funding agencies can define goal levels; it is flexible and adaptable after a tailoring.

In the EO domain it might be adopted to facilitate and improve CEOS WGISS Data Stewardship activities and achievements, though it needs to be adapted to take into account specific Earth Observation requirements and already existing Best Practices. Different mission datasets might have different targets and different maturity level ratings.

The DSIG analyzed two EO models: *Standards for Establishing Trusted Repositories for* ***USGS*** *Digital Assets* and *A Provenance Maturity Model from Environmental Software Systems. Infrastructures, Services and Applications book by* ***CSIRO***.

The DSIG has been working on a CEOS harmonized maturity matrix document, and she displayed the table of contents. They will use the Data Stewardship Maturity Matrix as starting point, analyze and collect other standards useful for the scope, and integrate and create a CEOS Harmonized Maturity Matrix. This will be followed by internal WGISS review and production of final version, and given to CEOS as a contribution. They plan a new review or approval at WGISS-43.

Richard asked if this matrix could be aligned to the data management principles. Iolanda replied that they made an analysis comparing the two, and this matrix is more comprehensive. The DMPs should match this matrix. It may be advisable to propose that the matrix be added to the document that the DMP-TF is currently working on.

Mirko said the idea was to use this in the WGISS Best Practices as it provides a target for mission management. It could be included in the WGISS BP, or could be tailored for DMP and submitted as a contribution. Iolanda said that it is difficult to decide which kind of activity to put in each level. What has already been done to categorize was already very difficult to decide. Mirko suggested working on the Maturity Matrix and deciding how it can be linked to the DMP document.

In response to a question from Richard, Mirko said that certification is for archives. Some of the points are already in line with ISO. The matrix authors decided that “by level” was the best classification; they referenced each with all the different standards.

Richard noted that the LTDP Guidelines need to be consistent with the matrix. Iolanda said that this is an activity they are already working on. Katrin strongly supported that the DMP needs to be aligned with the MM.

Dawn commented that some of points are subject to interpretation, so asked for more detail on each of the cells. Iolanda gave references for that.

## Report on Agency Stewardship Activities

### ESA - Heritage Data and Knowledge Preservation

Mirko Albani gave a presentation on preservation of heritage data and knowledge at ESA. He discussed the ESA Heritage Data Programme, and noted that ESA space science heritage data represent a unique, valuable, independent and strategic resource owned by ESA Member States. This is a mandatory program, provided it is funded at a sufficient level. He discussed the LTDP+ Earth Observation specific implementation activities.

Data and knowledge repatriation is occurring through a current migration project. One of the pillars of the ESA EO Ground Segment Evolution Strategy is the shift towards a service-based implementation of core ground segment functions. Mirko described the heritage EO assets recovery, consolidation and preservation. It begins with extracting data from archived tapes, identifying gaps to search for missing data and continues with an attempt to recovery it. The process continues with searching for all the documents, all the knowledge, going through boxes and boxes of documents and all the heritage media.

Andy asked about the data centers; Mirko replied that they are continuing to do their work but are no longer responsible for maintaining the archive.

Mirko discussed the Knowledge Management System (KMS) aimed at facilitating the tracking, management and preservation of the ESA EO data assets associated knowledge. He also discussed the new Heritage EO Assets Permanent Exhibition which has two areas: for each satellite, the hardware that was used, pictures, data, and other exhibits. For each milestone the technology evolution is shown in terms of hardware, media, catalog and photos.

Mirko listed the ESA/TPM missions in the Heritage Data Programme, as well as the data preservation evaluation criteria and the status for each mission, shown in readable matrix form. He also listed heritage datasets to be addressed in the coming months.

Kristi asked about lessons learned. A lot of the reason for this activity is due to changes in technology. ESA policy is to always keep the lowest level data, but the question should be asked is if there is anything that should be considered if one might predict the near future (20 years from today). One lessons learned is that documentation was not preserved, nor hardware. A few lessons are included in the best practice. Some examples of mistakes made in the past also help. Maintaining provenance will be essential to finding the master in the future.

Dawn asked if the hardware they are keeping, does it still work. Some does, some does not. The idea for the museum is education and display. Mirko added that some hardware can be kept if nothing else for spare parts.

Mirko concluded mentioning the ERS-1/2 mobile app that can visualize but also download data through DropBox. ATSR and RA L1 products were downloaded from ESA and a back-end process extracted specific variables to be displayed on the mobile device (through FTP with Mobile Product Server). The user can order the full L1 product through the mobile App; product retrieved from ESA (FTP) and downloaded in DropBox (Socket Protocol).

### NASA - Earth Science Data Stewardship

Dawn Lowe gave a presentation on NASA Earth science data stewardship. She described the objectives, project lifecycle events, general requirements for preserving bits, ensuring discoverability, accessibility, readability, understandability, usability, and reproducibility of results.

Dawn also discussed the Preservation Content Specification (PCS), covering eight categories of content: Preflight/pre-operations, science data products, science data product documentation, mission data calibration, science data product software, science data product algorithm input, science data product validation, science data software, and checklist. The content has to be collected from many different places; she described the scale of the effort.

Dawn concluded saying that NASA would like to see a broad international standard identifying preservation content – NASA’s PCS is a good starting point, as are ESA’s Long Term Data Preservation documents; ISO Technical Committee on Geographic Information/ Geomatics (TC 211) is working on a standard (ISO 19165).

Dawn summarized that NASA has been collecting Earth observation data from many sources for over 50 years; data and derived scientific products are a valuable asset requiring stewardship and preservation. Attention to preservation is needed throughout lifecycle; waiting for closeout phase of projects is too late. The PCS helps with planning ahead, and she would like to see an international standard on preservation content; they are looking for collaboration.

In response to questions, Dawn noted that they always have a backup on tape, and they test the tapes on a schedule; they try to anticipate the requests for the data to be ready for them, and are always looking ahead. They are using LTO just for backup because it is inexpensive.

Richard said that it is interesting to try to develop a standard, and recommended some groups that are working on this. In the frame of ISO there is a standard, and the [Consultative Committee for Space Data Systems (CCSDS)](http://www.ccsds.org/) is also involved.

**Action WGISS-42-16**: Mirko Albani and Dawn Lowe to check with CCSDS and ISO on the status of the development of a standard for data preservation.

## Data Preservation Discussion

Mirko suggested for discussion the topics of landing pages, glossary extension, transcription chains, and thesaurus.

DOI for knowledge: a persistent webpage which shows all of the relevant information about a data set. Landing pages are widely covered by the CEOS Persistent Identifiers Best Practices document v 1.1.

Open points:

1. Refer to the associated knowledge of a data set series via the landing page?
2. Dataset PID Landing page format – would it make sense to identify a minimum set of information which should be contained in there? Updating CEOS PID Best Practice with recommendations on landing page format and management?
3. WGISS#41: “Effort is underway for the new metadata repository to have a generator to automatically produce the landing page”. Is there any WGISS activity?

Katrin commented that they are generating landing pages from the collection metadata, and some recommendations would be useful. Andy said that NASA does have recommendations. Recommendation can be put in the PID BPs and NASA’s recommendation can be a starting point. Nitant said that they also have landing pages, and can contribute to the guidelines.

Andy added that a landing page is rendered from the CMR, and generated on demand. It is not a requirement, but it is available for anyone. It is an HTML rendering; NASA calls them Product Pages, which might be a better term for it, to distinguish from landing pages manually created by agencies.

Open points for PIDs for associated knowledge:

* Assign PIDs to data associated knowledge (e.g. documents)?
* Assign a PID only to the associated knowledge permanently archived?

Nitant commented that assigning more ids to additional components can become very complicated.

**Action WGISS-42-17**: DSIG team to identify what is available in the DIF and ISO for landing pages; to review what is recommended/required of NASA, ISRO and DLR for landing pages; and to make recommendations.

**Glossary Extension**

Points for Discussion:

* Should be extended to include Glossary of Science Keywords?
* Is there a need for harmonization with other organizations?
* Should some new identified terms be included? GEOSS Data-CORE, Open Data, Full data access.

Unless there is an urgent need, this activity should be tabled. Katrin said she likes the idea of linking it to other glossaries. This is easy, harmonization is much more difficult. DSIG agreed to follow up on linking the Glossary of Science Keywords to other glossaries.

**Transcription Chains**

Issue: Recovery of unique data still on heritage media via dedicated transcriptions or to fill identified gaps in master datasets is a recurrent need in many space agencies.

* Need to maintain heritage media transcription chains up and running: cost, difficulties
* Worth to coordinate and join efforts

Possible Activity for Discussion:

Coordination of set-up and maintenance of heritage media transcription chains at different organization for possible mutual support.

Possible Next Steps:

* Definition of “Transcription Chains” metadata to build an inventory; “Transcription Chains” information survey in CEOS agencies; Creation of a common inventory.
* Definition of further coordination and cooperation steps.

**Thesaurus**

For the moment semantic search with ontology is too complicated, and does not really serve any current need. The controlled list of science keywords is working for now. At CNES they have a thesaurus built 10 years ago; they need to decide whether to replace it, use the IDN, or build a new one?

Andy noted that the IDN has parameter-level controlled keywords; any new ones should be added to the IDN; Richard agreed that agencies can map the keywords to their local (language) keywords.

It was agreed that this activity can be closed.

# Technology Exploration Workshop on CLOUD COMPUTING

## Introduction

Satoko Miura introduced the Technology Exploration Interest Group Workshop on Cloud Computing, welcoming the participants and outlining the presentations for the workshop. She noted that WGISS will produce a report outlining the results of the workshop. Satoko also listed suggested questions for each of the reporting agencies to address.

## GS Evolution and EO Innovation Europe

Mirko Albani gave a presentation on GS evolution and EO innovation in Europe. ESA is addressing the questions of how EO data will be used in the near future, and how can Europe maximize the benefits on its public investments. He described the concept of having a network of exploitation platforms, where traditional data access (data pull, i.e. download, and data push) is to be gradually complemented by data hosted processing. The exploitation platform can also be depicted to comprise a combination of layers featuring specific functions and services.

EO Innovation Europe is a strategic partnership established between DG-GROW and ESA-EOP to ensure close technical coordination of research and development, prototyping and operational activities, and aligned programming and complementary funding. The common ultimate objectives are enabling large scale scientific and commercial exploitation of EO data, stimulating the innovation with EO data, maximizing the impact of European EO assets and preserving European independence.

Mirko listed current and upcoming activities at ESA, and concluded that since exploitation platforms are being developed at ESA and globally, the agencies need to ensure that they are coordinated and aligned.

## ESA Thematic Exploitation Platforms and Cloud Computing Activities

Sveinung Loekken gave presentation on a network of EO exploitation platforms. He began with the background and description of the evolving ground segment ops concept, and the exploitation platform concept and types. The thematic exploitation platforms (TEP) concept is a platform where all the necessary data and tools and services exist, and a plethora of exploitation platforms have been developed by agencies and international organizations.

TEPs are being planned in step with the emergence of cloud computing, virtualization, and hosted processing. Some specific objectives are to enable engagement, building capability, sustainability, and taking advantage of evolving technology. Exploitation platforms offer multiple advantages as they enable rapid data access, full focus on exploitation, synergistic use of data, community building, rapid prototyping, automated data processing framework, replicability of scientific results, cost-effective approach to scalable ICT resources, and development of new business/funding models.

Sveinung noted that ESA is starting at the pre-operations stage and should complete pre-processing by mid-2017. Some step enablers are under development and based on enhanced processing, but no developments as yet are linked to multi-sensor fusion or other capabilities. Several TEPs have concrete engagement with large stakeholder communities, and are already well-connected to complementary projects involving stakeholders. The goal is to create an architecture that is general enough, with interoperable capabilities, and open source. He listed technology and system and service status and capabilities, and listed some early adopters and results. He observed that interest has been very high and users are getting impatient – but need to slow down and manage the risk and expectations.

Next steps, from a programmatic context, are a common architecture and technology, enabling public sector benefits and industry growth, developing network of EO platforms, and adjusting evolving technical capabilities. There are significant opportunities for data and information service providers, clouds and ICT developers and service providers, science, applications, research institutes, platform developers and service providers, EO digital marketplace brokers, users, value adders, etc.

Sveinung concluded saying that TEPs are starting to see tangible results and provide necessary research and development, capabilities, and experience to address the expanded objectives of EOEP-5, in particular EO Innovation Europe and the Network of Platforms.

Martin asked what he sees as the main obstacles to TEPs. The real problems are programmatic and economic; they need an element of support from the public. Philippe Bally added that two departments of ESA are working together on this, but it is a prototype, not a model for operating.

For the process of selection of candidates, they have two profiles for selecting the winners. The team develops black boxes, where the chain is simple, and there are three families of users to see where they fit. Every few months a ranking exercise is performed, and this is not a simple algorithm. Andy requested a write-up of this; maybe at WGISS-43 there can be a session on user requirements.

## Cloud Computing and Security

Julien Airaud gave a presentation on security with cloud computing. He introduced the topic describing the special needs and issues associated with cloud computing security. He noted that the elasticity of the cloud applies to the security mechanisms; automation and orchestration provide an error free environment, and better risk management applies to the consumer and provider.

Julien noted that accidents (failure) or human errors are often underestimated, but systems should be built accordingly. He displayed a chart of Cloud Computing Vulnerability Incidents by cloud vendor.

Julien described some use cases of cloud security underway at several space agencies, and listed organizations that can provide guidance as agencies plan for this, such as the Cloud Security Alliance (CSA). Organizational security concerns include governance (contract), security responsibility (shared), and compliance to legal and sectorial regulations. Solutions can be found in the area of contracts, supplier assessments, compliance, audits of control, and risk management.

Julien described cloud computing risks, which include the usual information system security risks; although abstracted, the management plane is there, and is web or API based. Centralisation of everything owned is a concern, as are malicious insiders with high privileges from the back-office. The management interface is only as secure as users’ secret credentials, data can be intercepted in transit, and vulnerabilities apply to compute, network, and storage. A compromised node in a processing infrastructure can lead to data leakage, incorrect output, and infrastructure attacks.

Data processing solutions or architectures are built with performance in mind and security is left to the surrounding infrastructure. Data processing infrastructure is connected to untrusted resources (multi-tenancy in community/public deployment). One component can compromise the entire «cluster», and the provider cannot be trusted. The Virtual Machine or computing node is the new boundary of the system. So security must be integrated in the design pattern of the system. Infrastructure needs to be trusted; the code needs to be developed with multi-tenancy in mind the resources no longer needed should be killed. For data security, information architectures, data dispersion, information management lifecycles, and confidentiality should all be used. Data at rest and in transit should be encrypted; data erasing is the provider’s responsibility.

Access controls include using identity and access management, maintaining least privilege, creating suitable roles for users and multiple access keys and security groups, identity federation and SSO from internal sources, and trace actions. Security monitoring needs to be considered, as should be preparation and readiness for the coming incidents.

Nitant asked what the complexities are with multiple providers (multi-tenant cloud). Julien replied that in these cases it is best to encrypt, not trusting the provider; for multi-tenancy data will be lost so you can deploy a cloud access security broker. CNES ensures they are linked with DLP (data loss prevention).

Richard commented that the rules are quite complicated; is it contradictory to have a shared cloud with different rules from different countries? Mirko replied that this is where it is important to maximize and unify the effort, and to identify the contribution and role for the space agencies. The most difficult area can be the privacy rules, though Salvatore Pinto noted that from technical point of view it all depends on the type of data and type of policy; personal data does not have to be put on the cloud, and can be handled on the server side. Cloud providers are very good at security.

##  JAXA Approach on Virtualization and Cloud Computing

Satoko Miura presented the JAXA approach to virtualization and cloud computing. She began describing JAXA’s ground systems for EO satellites, and observed that they are always working toward cost reduction. To this end, she described a study initiated in 2010; Phase 1 target was to study the cost for system replacement, which system can work on the cloud, and can virtualization reduce the replacement cost. The study found many issues with placing servers on the cloud, and this option was discarded. They also considered virtualizing some systems with Hypervisor, but Hypervisor software is updated every three years

The Phase 2 target is to determine cost of operation and maintenance. This would involve a system migration toward a new “common” system, where servers are handled as “resource”; server procurement and application/system procurement can be divided, and service continuity would not be disrupted. However, failure on one server has impact on multiple functions; in the “micro” view performance may decrease, but in “macro” view, overall performance will be increased.

Satoko described target architectures for 2017 and 2018; the ideal future architecture has all the processing in a common supercomputer system. The current system has 17 servers running; the plan would reduce this number to eight. Another options is using a cloud on the premises, or super computers; discussions are ongoing. Some issues for using the cloud include cost for data download, network between cloud and JAXA system, varied license conditions, vendor lock-in, and security.

Another option under consideration is a user service on cloud, with an expansion of the user base among researchers, business and application users, and others.

Satoko added that benefits of the cloud, when used for data re-processing, mean that the data will be ready within much shorter period, with no need to maintain servers for occasional re-processing. If the data download cost issue and network issue is resolved, common data processing functions and data storage functions can be placed on cloud. User service portal may be the candidate if users are expected to increase drastically, or if “burst” type access is expected. Systems with almost fixed CPU/memory/storage requirement have a few advantages in the cloud.

JAXA’s Big Data challenges are related to data management and data analysis, as is data re-processing. Cloud computing may be a good solution for data analysis, but it is unclear if it will be for data management.

Richard commented that a disadvantage with the Google Earth Engine is that the vendor can do whatever they want with your code and data.

## ISRO Requirements and Research Issues for EO Data Processing Cloud

Nitant Dube reported that ISRO has decided to experiment on a private cloud, since moving to a public cloud can have significant bottlenecks. This will helps determine the requirements and considerations. He showed a roadmap of technology changes since 1960, beginning with visual interpretation and limited digital processing, and ending with cloud computing.

Infrastructure requirements for EO data processing include capability to migrate legacy and operational code to new infrastructure, without requirements for change in software. Reprocessing of old data requires huge computational loads, and EO data processing usually generates burst loads (LEO, GEO) and capability to optimally utilize the resources without addition of extra complexity in the software. Improved reliability in operations is necessary, as are time Series processing and analysis of EO data.

Cloud computing provides computing and resources (both hardware and software) on demand without worrying about the complexity and details of the underlying infrastructure. It also allows systems to scale-up and scale-down (both capacity and functionalities) based on requirements. Cloud allows users to create virtual organization, where resources can be optimally used, but performance, reliability, SLA, data control, standardization, and interoperability might be compromised.

Software requirements of satellite data processing are many: Capability to search service not only based on name or provider but also on parameters to meet the functional and quality requirements. A framework where new research can be used by existing applications and existing services can be used by new software is needed. Support for versioning in services is also a requirement; whenever services are enhanced and as new versions are released, there should be capability to use either the latest version of the service or fixed version based on user requirement. There is also a need for a Service Evaluator, which can verify the quality parameters based on data set submitted by the user. Ease of Use for service composition and option for the user to change the composed application are also requirements.

Nitant described a service oriented architecture, as well as a layered architecture, where software, platform and infrastructure are services.

EO cloud application scenarios include data/information products using multi-source EO data, providing access to dynamic, distributed and multi-source geospatial data for virtual reality applications. Another scenario is algorithm development and fine-tuning of products using multi-satellite data sources. Areas that need to be investigated are processing workflows, economy or scale, and interoperability.

Andy said that the suggestion of the EO cloud, where agencies combine their efforts is excellent.

## Computing in the Cloud at Geoscience Australia

Simon Oliver gave a presentation of cloud EO big data processing. He described GA’s cloud management capability, which consists of managed services, public cloud, and supporting infrastructure. He displayed infrastructure diagrams showing multi-disciplinary teams and public cloud services. Simon also described the National Computational Infrastructure (NCI) which provides their computational resources.

Simon observed that data downloading and analysis by many users has potential risks. Bringing scientists to the data can help mitigate these risks by ensuring everyone is working on the same data. To support this workflow, the NCI runs a Virtual Desktop Environment.

GA is using a range of cloud computing offerings to cover the range of requirements. They use AWS as an IaaS/PaaS provider and have built a continuous delivery pipeline to deploy systems into it. They also consume SaaS and PaaS services from other vendors and consume cloud IaaS from a specialist high performance computing eResearch environment with which they are partnered. He noted that they are still relatively immature with regards to governance but have quite a high level of understanding of the technical aspects, and are progressing rapidly.

GA is currently outsourcing in-house infrastructure to a managed service provider that provides a private cloud as part of their offering. They utilize cloud services at the National Computational Infrastructure (NCI) based at the ANU, and are making increasing use of AWS to host external facing web services. Applications are assessed for suitability against all of the providers at their disposal to ensure the most appropriate fit.

The ability to self-serve is one of the greatest benefits and this is in line with GA’s move to using multi-disciplinary teams to build their systems. The ability to experiment and try different things and then capture that configuration within scripts that can be reused is also valuable. Another other big benefit is the range of mature capabilities built into the platforms that are very stable, comprehensive and easy to use. Dual data centre deployments of applications can be as simple as clicking a few buttons.

Costs appear to be manageable for now; a general lack of governance – both technical and procedural – poses some potential threats and this is a focus going forward. It is imperative to establish safe patterns and architectures for our use, and have them followed.

Simon observed that cloud computing has a very important role in their projects and GA has positioned itself well with a range of providers to be able to ensure that the best location is chosen for each system, or component of a system. Some workloads are not well suited to public cloud but others are, and in fact the public cloud offerings provide benefits for these use cases far above what could be deployed locally.

Much of GA’s scientific data is large in volume, and while this is not the industry definition of “Big Data” it is nonetheless one of their biggest challenges. Curation and maintenance these ever-increasing datasets through their entire lifecycle, providing comprehensive data protection and archiving, as well as making it accessible to the people (and systems) that are interested in consuming it are all essential.

Cloud computing is potentially a solution to the big data challenge. Some of GA’s volumes are cost-prohibitive at the moment and part of the challenge is the desktop-centric way that many of the data are worked on by GA’s own staff. Many of GA’s largest datasets are held at the NCI and consumed there as well and this makes good sense, as moving large volumes around is still an impediment and a cost. The eResearch cloud environments may provide the most attractive solution to these challenges but other cloud services can help, and maybe complement this for data accessibility and distribution.

## Leveraging the Value of Data with Industry at NOAA

Martin Yapur began by acknowledging the many contributors of the NOAA Big Data Project, which deals with the accelerating user demand for NOAA data. The datasets are diverse, expensive and complex. There is concern that the data are underutilized due to accessibility issues. Martin displayed a conceptual overview of the project. He described the partnership announcement, which is an unusual, no-net-cost proposition, which received enthusiastic, cross-industry response; the need for further research and development became apparent. Five Cooperative Research and Development Agreements have been signed; collaborators are nuclei for data alliances and markets, and include members of industry, research, and academia.

Martin listed the partnership rules, and explained the methodology. The first data selected was NEXRAD, which is publicly available but difficult to use, and optimized for preservation*,* not access. The data is a highly popular dataset for use in industry. The utilization of the entire NEXRAD archive was never before realized, but it is now second in the US national observational value. New services are being referenced from NOAA NCEI’s website and a 60% decrease in archival data ordering has been observed. An improved NOAA archive is a direct and positive result of the AWS Big Data Partnership effort for NEXRAD Level II.

This partnership means that one valuable, large, unwieldy dataset has been “liberated” for wider use by industry and the public at no net cost, and new business opportunities are being created. New applications can be developed faster when data are co-located with processing. The next datasets to be used include geostationary satellite data, weather forecast models, fisheries bycatch information, and others.

Martin concluded that the Big Data Project success requires not just access to the data, but the expertise (algorithms, workflows, interpretive skill) and viable use cases. Potential developers and users encouraged to engage with BDP collaborators.

Nitant asked if there is any provision for students and researchers; Martin replied that those users are not the intent of the project. Richard asked what happens at the end of the contract; Martin replied that NOAA gets to choose what to do next; it is quite open.

## Cloud Hosting at USGS

Kristi Kline gave a presentation of cloud hosting at USGS. She mentioned that they are working on a large multi-agency contract with nine prime vendors and seven technical service areas. The USGS established a Virtual Data Center task for this project, in which all USGS websites are moving to the cloud. The potential for EO in the cloud includes access of Landsat data from cloud sources; cloud vendor and type of services is to be determined.

The advantages of cloud that Kristi listed include increased storage and processing resources; small projects and defined efforts are best served, and large projects show potential for improvements. She also detailed disadvantages, which include the cost of data egress, provenance/integrity of data in the cloud, potential procurement issues. Cloud systems still require high level of IT security as well as administration, engineering, software, the impact on the archive plan and on data management.

Andy asked if this is a cost savings, or is the primary goal making the data more accessible; Kristi replied that it could be a cost saving if the vendors provide the data for free, and only charge to for the services they provide. Nitant asked about the service level agreement; Kristi replied that the intent is to have an agreement for the services they provide.

Brian commented that using various cloud providers may cause more problems; if all the data is hosted by a single cloud provider it would be much simpler. He added that with large datasets the data has to be in close proximity to the processing system. Kristi noted that they would likely use multiple cloud vendors, but are not yet clear on the selection process.

## Assessing Applications of Cloud Computing to NASA’s EOSDIS

Chris Lynnes\* gave a presentation assessing applications of cloud computing to NASA’s Observing System Data and Information System (EOSDIS). He began with a description of the overall approach of developing prototypes with a focus on public clouds, and leveraging existing software. He described prototypes for data archives.

Chris indicated that one solution is to move more analysis closer to the data, and described cloud analytics prototypes, and the advantages of such a paradigm shift, where scientists will work on data in place instead of downloading, with high-value data in databases, and pre-existing toolsets easy to find and use. He also described application-hosting and processing prototypes, where off-the-shelf systems reduce effort for hardware procurement and deployment, and testing, deployment, scaling, and failover are automated. In this scenario, science users get new capabilities sooner, and NASA gets more reuse and lower hardware costs.

The current status of EOSDIS Cloud Progress is:

* Archive Prototypes: Serving data to the public from Alaska Satellite Facility Web Object Storage prototype in AWS Simple Scalable Storage; end-to-end lambda workflow has been demonstrated for the ingest/archive management prototype. Global Imagery Browse Service is undergoing system testing.
* Analytics Prototypes: NEXUS analytics algorithms benchmarked vs. Giovanni. Roughly order-of-magnitude speedup is achieved on a cluster; now porting to cloud-native architecture.
* Application Hosting Prototypes: NASA-Compliant General Application Platform (NGAP) is now authorized to operate publicly. Earthdata Search client is operational and accessible to the public in AWS, and modified processes account for new costing mechanisms.

However, recognized hurdles include vendor lock-in, future storage costs, uncapped egress costs, and security restrictions and network trust.

Kristi asked about their plans for data and provenance integrity; Chris replied that this will come into play with reorganizing the data so that it can be useful. Kristi also asked if they are pushing the newest data into the cloud. Chris replied that they are, for the prototypes.

## Data Cube Use of Cloud Computing by CEOS

Brian Killough\* gave a presentation of the CEOS data cube use of cloud computing. He began with the vision for the use of cloud computing in the CEOS data cube in terms of deployment, hosting, and processing. He explained that they use Amazon Web Services (AWS) and are comparing ingestion times using AWS data and user uploaded data. They are testing “on-demand” and “spot” processing instances (EC2) to optimize operations and costs, and also testing remote connections to a data cube using an API to allow QGIS or ArcGIS functions.

Google Earth Engine has several advantages, such as free, open, global datasets, powerful analysis tools for JavaScript and Python, and datasets are growing daily. However, using it builds a commercial dependency, has limited time and spatial scales, provides cloud-based computing only, Google “owns” all of the data, and missing datasets are a problem. Brian commented that data available from Google and Amazon is mostly Level-1 data, but some recent efforts are underway to provide pre-processed Level-2 products by Google (Landsat and Sentinel-1).

Brian presented a comparison of results of Data Cube Water Detection and Google Earth Engine Aqua Monitor. The results are almost the same, but in GE you cannot get back the pixel level data; it provides rapid results, but complexity and detail missing.

Brian raised the following questions for WGISS:

* The data cube project is planning to develop a Future Data Architecture (FDA) prototype project where data cubes will be tested. This project would certainly test cloud-computing. How might WGISS get involved?
* Does WGISS have any specific advice or thoughts on the use of cloud-based resources to support data cube deployment, storage and processing?
* Does WGISS know of any other cloud-based computing resources that may be interested in providing free credits for testing the CEOS data cube architecture?

## Cloud Computing Discussion

Satoko reviewed the answers that each agency gave to questions posed to the presenters. She noted that the various agencies are in various stages of cloud computing; some are in study stages, some are in pre-operational stages, and some are operational, but that all agencies seem to be going in the same direction. Satoko noted the following key points raised during the session:

* Move User activities to the Data (ESA)
* Move more analysis closer to the data (NASA)
* Cloud Interoperability (ISRO)
* Big Data Project success requires expertise (algorithms, workflows, interpretive skill) and viable Use Cases (NOAA)

Satoko requested discussion on the topics of cloud security, EO cloud, and private clouds. Andy added that WGISS will be providing much of the information gathered as a contribution to the FDA report, and this contribution should be valuable. Over the next few weeks a summary will be developed and reviewed by WGISS.

The following points were raised, or comments made:

Kristi: Understanding where the assets are in the cloud. The concept of ARD. WGISS can work on how best to make ARD visible, accessible.

Brian: WGISS could begin to compile lessons learned, advantages/disadvantages of cloud computing. WGISS could compile a list of major large datasets that now exist in the cloud.

Lubia: To focus on the analysis, emphasizing the need to determine the usability of the data in the cloud.

Nitant: How to develop cloud-aware (or cloud-ready) software. This paradigm reduces the complexities to the users by pushing the data to the cloud.

Cristiano Lopes: To get industry and science users more involved in the paradigm shift, where many validations are already done, toolboxes are provided, and processing is maturing.

Chris: How to maintain interoperability when the services are being combined with the data.

Richard: What interoperability is foreseen? He also raised the economic issue of cloud computing. Andy commented that for some agency applications cloud computing saves money; for others it does not. For example, NASA is saving money by putting the CMR in the cloud, but for many other applications it does not save money. Mirko noted that ESA sees some cost savings for what they have moved to the cloud. Richard added that some datasets are interesting to cloud providers, while others, such as heritage data, are not. Martin commented that there is potential for a cost-savings path, but the cloud providers are primarily interested in profit. Kristi added that paying for egress from the cloud is too expensive, but if the cloud providers provide the data for free it may be less costly; without limits on the data there would be an explosion of utilization.

# WGISS Plenary, Part II

## Future Webinar Discussion

Shinichi Sekioka gave a presentation to introduce and discuss the topic of the WGISS Technology Exploration group developing webinars for public use. There are two target audiences: For the general or beginner listener, an overview of the topic, about 30 minutes long. For the expert listener, an overview of the topic, and an additional 30 minutes of technical discussion.

For the first webinar, the Technology Exploration group invited Andy, the WGISS chair, to be the first speaker. Two potential topics are federated user identity, and interoperable metadata models.

A CEOS Wiki is also proposed, where the webinar logistics will be provided and advertised.

The topics proposed at WGISS-41 are also still pertinent:

* Big Data, HPC and Cloud Computing: CEOS needs and challenges, distributed data centers, data processing (including data cube), data distribution, data analysis, API and use of standards, network (bandwidth, application).
* Searching for free satellite data from CEOS agencies
* GCMD/IDN Keywords – what are they, how to add to these lists
* Search relevancy for collection searches
* Data quality semantics and augmentation of metadata
* Visualization of data: web-based visualization, volume rendering, tiling, augmented reality
* Using authentication/SSO
* Metrics of usage, metrics of datasets
* Crowd Sourcing

Martin volunteered WISP to help set up these webinars. Andy suggested a general audience first, and cloud computing for the topic. Yonsook added that WGCapD has suggested one on how to search for agency data, and she is willing to coordinate that. Mirko noted that these approaches to provide and submit a short report with recommendations is good.

**Action WGISS-42-21:** WISP team and Kristi Kline to research technology requirements for a CEOS wiki.

**Action WGISS-42-22:** Technology Exploration team to send an email to WGISS-All asking for volunteers to speak on specific topics for Technology Exploration webinars.

**Action WGISS-42-23:** Andy Mitchell and Mirko Albani to discuss suggested webinar topics with WGCapD chair.

## Future Meetings

Mirko Albani listed the following planned and projected upcoming WGISS meetings:

WGISS-43: April 3-7, 2017, in Annapolis, Maryland, USA, hosted by NASA.

WGISS-44: September 2017 in Asia; in discussion with ISRO, ROSCOSMOS, GISTDA, ANGKASA, VAST.

WGISS-45: March 2018 in Southern Hemisphere; SANSA, CSIRO/GA, INPE, CONAE under consideration; Martin suggested facilitating with Colombia; WGISS Exec will inquire about rules around non-member hosting.

WGISS-46: September 2018, in Europe: DLR, CNES, UKSA, and GSDI under consideration.

WGISS-47: March 2019, North America.

Andy described arrangements for WGISS-43. He announced that the meeting would be hosted by NASA Goddard Space Flight Center and held at the Historic Inns Annapolis (58 State Circle, Annapolis, Maryland 21401 USA). The dates are April 2nd to 7th 2017, and lodging is also blocked at the same inn. Andy listed things to do in Annapolis, adding that he is hoping to have a tour of the visitor center and the satellite lab at Goddard Space Flight Center. Three international airports serve the area.

Andy also gave an overview of the potential agenda.

**Action WGISS-42-24:** Mirko Albani and Andy Mitchell to speak with the CEO about guidelines regarding hosts for WGISS meetings.

## Chair Summary

Andy gave a summary of the meeting, beginning with a display of the group photo. He discussed the two GEO foundational tasks that are of interest to WGISS (GD02 and GD07), and outlined proposals for further GEO work. WGISS identified a new system level team to work with the GEO Sec, GEOSS Portal and DAB team toward improving CEOS agency assets discoverable and accessible through IDN, CWIC and FedEO. Andy also mentioned the upcoming GEO Plenary, and the 1st Virtual Workshop GEO DAB API’s.

Andy noted the SEO request for continued support to expand the connections from mission archives to the COVE tool; future targets include Sentinel-2 and CBERS-4. Additionally SEO requests WGISS’ help to find an approach for automated discovery, processing, downloading and ingesting of data to support users with data cubes.

The CGMS Global Data Dissemination (WG IV) has many tasks in common with WGISS; WGISS can and should support action items of WG IV to enable interoperability, and invite WG IV members to WGISS-43.

Andy listed several potential technology exploration topics; WGISS will develop webinars on such topics. The first webinar is planned for December 2016

* Discovering and accessing data in future architectures
* Geospatial applications workshop on the use of CEOS data
* Metadata interoperability
* User authentication interoperability
* Hack-a-thon on client development (or API use) to access CEOS data
* Continued discussion of Cloud Computing

Andy noted that WGISS has been asked to review the latest FDA Report, and listed a number of pertinent comments made during discussion that can be presented to the FDA team. WGISS will create a summary report of the Cloud Computing Workshop for this purpose. Areas of cloud computing for WGISS to research include: will cloud save the agencies money? What interoperability issues will cloud solve? How best to make ARD discoverable/accessible. What cloud enabled applications should be pursued; interoperability complexity with the use of multiple cloud vendors. Cloud-aware software.

Andy displayed a diagram for consolidation of current CWIC/FedEO/IDN overall architecture to quickly address some of the identified open issues.

The Data Stewardship Interest Group presented the Scientific Data Stewardship Maturity Matrix; this matrix can be aligned with the Data Management Principles and the LTDP Recommendations.

The documents adopted during WGISS-42 are:

* CEOS OpenSearch Best Practices
* Data Purge Alert
* Associated Knowledge Preservation Best Practices
* CEOS Maturity Matrix

## WGISS-42 **Actions**

The actions resulting from WGISS-42 are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number** | **Category Category** | **Action** | **Actionee** | **Due Date** |
| **WGISS-42-01** | GEO | Andy Mitchell and Mirko Albani to recommend to the GEO-Sec (Osamu Ochiai), if GD-07 becomes an initiative in the GEO Work Plan 2016-18, to move the Data Management Guidelines Task to GD-02. | Andy Mitchell, Mirko Albani | Oct-31-2016 |
| **WGISS-42-02** | Carbon | Andy Mitchell and Mirko Albani to obtain from WGClimate the final results of the ECV Inventory Gap Analysis for Carbon. | Andy Mitchell, Mirko Albani |  |
| **WGISS-42-03** | Carbon | Ken McDonald to research the GEO Carbon Portal. | Ken McDonald |  |
| **WGISS-42-04** | Carbon | Andy Mitchell, Mirko Albani, Martin Yapur, and Ken McDonald to define the requirements for a CEOS Carbon Portal, working with WGClimate and the Carbon action coordinator, Mark Dowell. | Andy Mitchell, Mirko Albani, Martin Yapur, Ken McDonald |  |
| **WGISS-42-05** | Data Access | Yves Coene to distribute to WGISS the new OGC Specification for OpenSearch. | Yves Coene | Oct-06-2016 |
| **WGISS-42-06** | Data Access | OpenSearch team to finalize and post the CEOS Open Search Best Practice after making minor editorial changes referencing the OGC. | OpenSearch team | Oct-31-2016 |
| **WGISS-42-07** | Data Access | OpenSearch team to circulate the Open Search Developer’s Guide after allowing one additional month for comments. | OpenSearch team | Oct-31-2016 |
| **WGISS-42-08** | Data Access | Mirko Albani and Olivier Barois to research using the GEO-JSON encoding, and identify members that can participate in the OGC group. | Mirko Albani, Olivier Barois |  |
| **WGISS-42-09** | Data Access | WGISS representatives to volunteer (by sending email to Andy Mitchell and Mirko Albani) for the WGISS Connected Data Assets System Level Team. | WGISS Representatives | Oct-06-2016 |
| **WGISS-42-10** | Data Access | Yonsook Enloe to initiate a teleconference for the first meeting of the WGISS Connected Assets System Level Team. | Yonsook Enloe | Oct-31-2016 |
| **WGISS-42-11** | Data Access | Yonsook Enloe to revise the GCI User Requirements document to insert draft requirements addressing the problem of harvested/cached datasets and duplicate datasets in GEOSS.  Initiate a review of the new draft GCI User Requirements with members of WGISS. | Yonsook Enloe |  |
| **WGISS-42-12** | Data Access | WGISS to explore to possibility of including publishing services related to EO datasets in the IDN. | WGISS representatives |  |
| **WGISS-42-13** | Data Access | Andy Mitchell, Yonsook Enloe and Simon Cantrell to work with WIGOS and CGMS metadata representatives to identify common collection-level metadata elements between the IDN, CGMS, and WIGOS. | Andy Mitchell, Yonsook Enloe, Simon Cantrell |  |
| **WGISS-42-14** | Data Use | Gábor Remetey to coordinate (via WGISS) GSDI contributors to a future WGISS geospatial applications workshop on the use of CEOS data. | Gábor Remetey |  |
| **WGISS-42-15** | Data Stewardship | DSIG team to consider developing a video on the value of WGISS Data Preservation activities for the Faces of CEOS series. | DSIG team |  |
| **WGISS-42-16** | Data Stewardship | Mirko Albani and Dawn Lowe to check with CCSDS and ISO on the status of the development of a standard for data preservation. | Mirko Albani, Dawn Lowe |  |
| **WGISS-42-17** | Data Stewardship | DSIG team to identify what is available in the DIF and ISO for landing pages; to review what is recommended/required of NASA, ISRO and DLR for landing pages; and to make recommendations. | DSIG team |  |
| **WGISS-42-18** | **W** Technology **GISS-42-02** | WGISS-42 participants to review Satoko Miura’s outline of the Cloud Computing Workshop and provide feedback. | WGISS-42 Participants | Oct-06-2016 |
| **WGISS-42-19** | Technology | Technology Exploration team to create a summary report of the Cloud Computing Workshop; the summary should be tailored for input to the FDA report. | Technology Exploration team | Oct-31-2016 |
| **WGISS-42-20** | **W** Technology **GISS-42-02** | Technology Exploration team to investigate user authentication interoperability. | Technology Exploration team |  |
| **WGISS-42-21** | WGISS Webinars | WISP team and Kristi Kline to research technology requirements for a CEOS wiki. | WISP team |  |
| **WGISS-42-22** | WGISS Webinars | Technology Exploration team to send an email to WGISS-All asking for volunteers to speak on specific topics for Technology Exploration webinars. | Technology Exploration team |  |
| **WGISS-42-23** | WGISS Webinars | Andy Mitchell and Mirko Albani to discuss suggested webinar topics with WGCapD chair. | Andy Mitchell, Mirko Albani |  |
| **WGISS-42-24** | WGISS Logistical Support | Mirko Albani and Andy Mitchell to speak with the CEO about guidelines regarding hosts for WGISS meetings. | Mirko Albani, Andy Mitchell | Oct-31-2016 |
| **WGISS-42-25** | WGISS Logistical Support | WISP team to compile a mailing list of members who regularly attend WGISS meetings for specific communications. | WISP team | Oct-31-2016 |

## Adjourn

Andy adjourned the meeting, thanking the participants for their contribution to a productive and successful meeting. He also thanked ESA/ESRIN for their wonderful hosting. WGISS looks forward to their support, as there is an explosion of work for the future that is very exciting and will prove to be very useful.

# Glossary of Acronyms

API Application Programming Interface

ARD Analysis Ready Data

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

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

IDN International Directory Network

ISO International Standards Organisation

LSI Land Surface Imaging

LTO Linear Tape-Open

NRT Near real-time

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

TOA Top of the Atmosphere

ToR Terms of Reference

UML Unified Modelling Language

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