MINUTES

OF THE

38th MEETING

OF THE

CEOS WORKING GROUP ON
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Moscow, Russia

29 September to 2 October, 2014

Hosted by

Roscosmos

Research Center for Earth Operative Monitoring

Joint Stock Company “Russian Space Systems”

Table of Contents

[1 WGISS Plenary Session, Part I 5](#_Toc403329834)

[1.1 WGISS Welcome, Introductions, Adoption of Agenda 5](#_Toc403329835)

[1.2 Logistics Information 5](#_Toc403329836)

[1.3 Host Welcome 5](#_Toc403329837)

[1.4 WGISS Infrastructure Support Project (WISP) 5](#_Toc403329838)

[1.5 WGISS Chair Report 6](#_Toc403329839)

[1.6 CEOS Executive Officer (CEO) Report 6](#_Toc403329840)

[1.7 Systems Engineering Office (SEO) Report 7](#_Toc403329841)

[1.8 GEO Secretariat Report 8](#_Toc403329842)

[1.9 GEO Infrastructure Implementation Board (IIB) Activities 8](#_Toc403329843)

[1.10 WGISS Terms of Reference 9](#_Toc403329844)

[1.11 Future Meetings 10](#_Toc403329845)

[2 Agency and Liaison Reports 11](#_Toc403329846)

[2.1 Academy of Opto-Electronics (AOE) 11](#_Toc403329847)

[2.2 European Space Agency 11](#_Toc403329848)

[2.3 Geoscience Australia 11](#_Toc403329849)

[2.4 Global Spatial Data Infrastructure (GSDI) Association 11](#_Toc403329850)

[2.5 Japan Aerospace Exploration Agency (JAXA) 12](#_Toc403329851)

[2.6 National Aeronautics and Space Administration (NASA) 12](#_Toc403329852)

[2.7 National Oceanic and Atmospheric Administration (NOAA) 12](#_Toc403329853)

[3 WGISS Projects 13](#_Toc403329854)

[3.1 CEOS Water Portal Project 13](#_Toc403329855)

[3.2 Disasters Recovery Observatory (RO) Project 13](#_Toc403329856)

[3.2.1 Introduction 13](#_Toc403329857)

[3.2.2 RO Presentation to WGDisasters 13](#_Toc403329858)

[3.2.3 Recovery Observatory Version 1.0 14](#_Toc403329859)

[3.2.4 Discussion 14](#_Toc403329860)

[4 Interest Groups 16](#_Toc403329861)

[4.1 Technology Exploration Interest Group 16](#_Toc403329862)

[4.1.1 CNES - Big Data Technologies (Hadoop) 16](#_Toc403329863)

[4.1.2 Geoscience Australia Data Cube 16](#_Toc403329864)

[4.1.3 Big Data and Cloud Computing for CEOS 17](#_Toc403329865)

[4.1.4 Big Earth Data Initiative (BEDI) 17](#_Toc403329866)

[4.1.5 TEIG White Paper 17](#_Toc403329867)

[4.1.6 CEOS Plenary Side Meeting 18](#_Toc403329868)

[4.2 Data Stewardship Interest Group 18](#_Toc403329869)

[4.2.1 DSIG Plan of Activities 18](#_Toc403329870)

[4.2.2 DSIG Data Stewardship Best Practices/Guidelines 18](#_Toc403329871)

[4.2.3 PV2013 19](#_Toc403329872)

[4.2.4 NASA Data Stewardship Activities 19](#_Toc403329873)

[4.2.5 USGS Data Stewardship Activities 20](#_Toc403329874)

[4.2.6 JAXA Persistent Identifiers Study and Discussion 20](#_Toc403329875)

[4.2.7 Persistent Identifiers Experience within LTDP WG 20](#_Toc403329876)

[4.2.8 Purge Alert Procedure and List 21](#_Toc403329877)

[4.3 CEOS Interoperability Interest Group 22](#_Toc403329878)

[4.3.1 International Directory Network 22](#_Toc403329879)

[4.3.2 Federated Earth Observation (FedEO) 22](#_Toc403329880)

[4.3.3 CEOS WGISS Integrated Catalog (CWIC) 23](#_Toc403329881)

[4.3.4 CEOS OpenSearch Project 26](#_Toc403329882)

[5 CEOS OpenSearch Workshop 28](#_Toc403329883)

[5.1 OpenSearch Characteristics 28](#_Toc403329884)

[5.2 CEOS OpenSearch Best Practice 28](#_Toc403329885)

[5.3 OpenSearch Implementations and Demonstrations 28](#_Toc403329886)

[5.3.1 IDN/CWIC 28](#_Toc403329887)

[5.3.2 GCMD/CWIC 29](#_Toc403329888)

[5.3.3 FedEO 29](#_Toc403329889)

[5.3.4 CNES 29](#_Toc403329890)

[6 WGISS Plenary Session, Part II 30](#_Toc403329891)

[6.1 WGISS Chair Summary 30](#_Toc403329892)

[6.2 Adjourn 31](#_Toc403329893)

[6.3 WGISS-37 Actions 32](#_Toc403329894)

[6.4 Adjourn 32](#_Toc403329895)

[7 Glossary of Acronyms 34](#_Toc403329896)

List of Participants

CAS-AOE Chaoliang Wang, Guangyu Liu\*

CEOS-CEO Kerry Sawyer\*

CEOS-SEO Brian Killough\*

CNES Richard Moreno (WGISS Chair), Jérôme Gasperi, Mathieu Gond, Laurent Baresse\*,

 Benoit Frezouls\*, Michelle Piepgrass (WGISS Secretariat)

DLR Katrin Molch\*

ESA Mirko Albani, Yves Coene, Stefano Nativi\*, Mattia Santoro\*

EUMETSAT Michael Schick, Uwe Voges\*

GEO Secretariat Osamu Ochiai\*

Geoscience Australia Jonathon Ross

GSDI/HUNAGI Gábor Remetey-Fülöpp

INPE Lubia Vinhas\*

ISRO Nitant Dube

JAXA Satoko Miura, Yoshiyuki Kudo, Shinichi Sekioka, Hitoshi Tsuruma

NASA Andrew (Andy) Mitchell (WGISS Vice-chair), Thomas Cherry, Yonsook Enloe, Lingjun Kang\*,

 Michael Morahan, Doug Newman, Thomas Plumb, Archie Warnock\*

NOAA Martin Yapur, Ken McDonald\*

NRSCC Dingsheng Liu

Russian Space Systems Valery Zaichko, Tamara Ganina, Alexey Gladkov, Ovnan Tokhiyan, Marina Krasnych, Kirill Emelyanov, Anna Prokopchic, Elana Maroko

UKSA Wyn Cudlip

USGS Kristi Kline\*

\* Via web conference or email

# WGISS Plenary Session, Part I

## WGISS Welcome, Introductions, Adoption of Agenda

Andy Mitchell (WGISS vice-chair) welcomed the participants to WGISS-38, explaining that Richard Moreno (WGISS chair) was unavoidably delayed one day.

Andy thanked Tamara Ganina for all the excellent arrangements for the meeting. He asked those present to introduce themselves. He reviewed the agenda and it was adopted with no significant modifications.

## Logistics Information

Tamara Ganina described the logistics of the meeting for lunch, breaks, group photo, tour of Roscosmos facility, hosted dinner, and riverboat tour. She also made a few suggestions for sightseeing, and provided emergency contact information.

## Host Welcome

Valery Zaichko, deputy head of Roscosmos Department for automatic space complexes and systems, greeted the participants to WGISS-38 as colleagues and friends, hoping that everyone would have an enjoyable visit to Moscow, and thanking everyone for traveling to Russia. He noted that it is pleasant to greet friends from previous meetings. Valery said that it is his hope that this work in Moscow will lead to more intensive cooperation between CEOS WGISS and Roscosmos, and on behalf of the Russian space agency he desired fruitful work in Moscow for WGISS, especially toward promoting integration of Russian EO data to the international community.

Valery also desired the participants to enjoy the highlights of Moscow and of the warm and welcoming Russian people. To this end his office has arranged a dinner hosted by the space agency, and a riverboat dinner cruise. His office has also arranged a visit to the EO data processing center, where participants will be able to observe first-hand the services they provide and the issues they face in terms of cooperation with the international community.

Roscosmos has joined the CEOS Charter for disaster recovery as the 15th member; their efforts for disaster response within the CEOS framework, with WGDisasters, and with WGISS will be evident during the tour of the center. Roscosmos is reaching a point of standardization for EO data, and would like to consult with WGISS on this. He addressed the colleagues from China, noting an upcoming conference in Hainan, and with the help and support of the Chinese colleagues he hoped many can join. Roscosmos will also participate at the CEOS Plenary in Norway.

The Joint Stock Company “Russian Space Systems” was specially founded by Roscosmos in 1999 and designed for performing the duties on operation of the Russian space remote sensing facilities. Russian Space Systems is an operator of the Russian space remote sensing systems, the Roscosmos information center of remote sensing, the Roscosmos Center in the Unified System of Information on the World Ocean Condition (ESIMO), and the Roscosmos-authorized representing organization in the International Charter on Space and Major Disasters.

Valery displayed a diagram of the Russian EO missions of 2006-15, and described their EO data receiving and recording facilities. Operator of Space Remote Sensing Systems provides customers with remote sensing data in internationally accepted processing levels and standards include Level 0, 1, 2A, 2B, and 3, and standard data products.

Since 2010, under the Roscosmos supervision the sub-satellite validation observation system has been established aimed first of all at evaluating geometric and radiometric characteristics of high and ultra-high resolution data. There is a need for an interdepartmental system of remote sensing data product validation in terms of their full thematic use for solving specific tasks in accordance with the claimed characteristics. The system being established will provide validation by the main characteristics of target equipment of all existing and future Russian remote sensing space facilities of visible, near infrared, and radar regions.

Valery described the geoinformation services of the Russian Space Remote Sensing Systems Operator and its evolution, and also its participation with the International Charter on Space and Major Disasters.

RSS is conduction joint research in the field of hyperspectral Earth observation for social-economic tasks (natural resources, ecology, emergency etc.), using real-time data received from Russian satellite RESURS-P1/2 and ALOS3 test data. The main target is data quality methods and algorithms of hyperspectral data products validation. Global Earth observation data (optical & SAR) exchange between Russian and Japan space systems is ongoing, and ALOS2/3 data distribution in Russia (governmental sector) is using facilities of the Russian space remote sensing Operator (receiving stations, data processing, thematic applications, geoservices).

Andy Mitchell commented that the work Roscosmos is doing is exciting, particularly with the Geoportal and integration with CWIC. He was also glad to hear that they have joined the Charter, and of their efforts for Disaster Recovery and for standardization. Yonsook added that the CWIC team has really enjoyed working with Tamara, Alexey and Ovnan and are looking forward to learning which data will be available to the global community via CWIC.

Valery wished the participants a fruitful day and a very productive week.

## WGISS Infrastructure Support Project (WISP)

Martin Yapur gave the report of the WGISS Infrastructure Support Project. He listed the members of the team, and displayed a diagram of the latest WGISS structure. He gave connection details for remote access to the meeting, and explained how presentations would be stored on the Google Drive account (WGISS.support@gmail.com). Past presentations can be requested from Anne Kennerley (anne.kennerley@noaa.gov).

Andy continued with information on changes to WGISS website. Changes were recommended at WGISS-37, and the Technology Exploration Interest Group, which is a forum for exchange of technical information and lessons-learned, agreed to initiate upgrades that include ways to facilitate the exchange of information and documents. A few changes will be recommended for the cover page, but key improvements are the addition of two links on the left hand side of the WGISS page for links to new pages on WGISS Data Tools and Services, and Open Source Technology. The new pages will include a table with name, agency, categorization, data handling, geolocation, data visualization, and if it is open source. There will also be a hyperlink to the actual tool website. Questions or comments should be addressed to Andy Mitchell and Satoko Miura.

The CEOS webpage is also being redesigned – it is six years old, with a lot of content, and 35,000 visitors in the last year. It needs an architectural and aesthetic update. CEOS is considering a document management system, an “action” tracking tool, meeting registration, and social media capabilities. The final display will presented for review at the CEOS Plenary.

In response to a question from Wyn Cudlip, Andy said the plan includes archiving historical documents. Martin added that all the presentations since WISP began at WGISS-27 can be included.

**Action WGISS-38-8**: Interest groups and projects to update their pages on the WGISS website.

## WGISS Chair Report

Richard Moreno gave the WGISS Chair report. The ‘WGISS future’ discussion at WGISS-37 highlighted the need for concrete and visible/tangible output. A list of such current tangibles are the OpenSearch Best Practices document, and alignment of the various OpenSearch implementations. Other tangible outcomes are the efforts of the Data Stewardship Interest Group and concrete output from LTDP, and the Recovery Observatory (RO) Project, with infrastructure support of WG Disasters. Another concrete output is the Emerging Technologies White Paper, and the related side meeting planned for the CEOS Plenary. The IDN is the most important contributor to GEO, and is continuing its strong evolution, as is the interoperability between CWIC, FedEO, and other agencies. At SIT several statements have been made portraying a positive image for CWIC.

In terms of external engagements, WGISS has received requests from SEO, WGClimate, WGCapD, GEO, and the Precipitation VC.

Richard also presented information on the Data Management Principles Task Force. He explained that at the GEO-X Plenary the establishment of a Data Management Task Force (DMP-TF) was endorsed to identify data management principles. The DMP-TF kicked off on 16 May with Terms of Reference (ToR) drafted by the Infrastructure Implementation Board (IIB) and Data Sharing Working Group (DSWG), with 19 individuals nominated by 16 GEO Members and Participating Organizations, and administrative support provided by the GEO Secretariat. The task force was asked to produce a report for consultation aiming to move steps along agreed principles to be adopted in 2015 together with the new implementation plan. The TF has a very short period to collect and review best practices and procedures, identify and describe categories for procedures, and assess feasibility of principles, and produce draft and final reports for the Plenary.

The value of each EO is maximized through data life-cycle management, including the following foundational elements: discoverability, accessibility, usability, preservation, and curation. Principles for each of these elements were listed.

Mirko commented that it is not clear is how the principles will be implemented. Some of the discussions includes the best way to ensure the providers apply these principles. The discussion will be ongoing and a set of documents will be developed to make sure that the best practices are used and to ensure the principles are applied. Wyn added that there is still the issue of harmonization of the quality indicators. The point of view of WGCV is the quality indicators are not in line; they are working on it, but for the moment it is not possible – more a wish than a reality. Mirko agreed that this discussion should continue, and it is a very complex area. Nitant commented that meteorological operational agencies calibrate their products to the same reference, and publish their results on their websites. This is a valid and key approach due to the nature of the international data exchange. But that is only one domain of EO data.

Richard described the GEO task IN-02 (Earth datasets). He also described GEO component: IN-02-C1-Advances in life-cycle data management, whose task sheet can be found at <http://www.earthobservations.org/ts.php?id=135>. Richard is the PoC.

Mirko is participating in the IIB as the ESA representative. Mirko will be making a presentation, and will take time to discuss what would be the contribution from WGISS. The list of outputs and activities of IN-02-C1 is quite large. There was discussion on the coordination of the task sheet, and Satoko said they are working on the communication and structure since it is not working so well. Andy asked how WGISS could be assigned to a task, noting that the QA4EO is was part of IN-02-C1. Satish (WGCV Chair) was asked to be PoC, but when he realized that 90% of the work is more related to WGISS than to WGCV, he asked WGISS to take it. The GEO secretariat is responsible for the list of participants, and also for the work plan.

Richard also discussed the SIT workshop 2014 side meeting for CEOS working groups and virtual constellations. Richard noted that a topic of interest to WGISS was on developing working group interactions. There was also a session on data management and access where Richard gave a presentation, but the other presentations were more of interest to WG Disasters. Andy noted that he spoke with Steve Neeck about the best way for WGISS to interact with the VC, but Steve turned the question around, saying that he prefers if the WG comes to him with what they can offer. Yonsook commented that the meaning of “data access” is quite different in different communities, and that changing approaches may be a positive thing. A good starting point may be to ask them what data they want, and then trying different approaches to give it to them.

Wyn asked about the longevity of LSI VC; Jonathon said that it is not being coordinated, and a set of options will be presented at the CEOS Plenary regarding how to proceed with the VC.

**Action WGISS-38-2:** Martin and Andy to ensure a GoToMeeting is scheduled and publicized to WGISS-38 participants for the side meeting at the Plenary.

**Action WGISS-38-3**: WGISS Exec to review the GEO task sheets to identify areas where WGISS can contribute.

## CEOS Executive Officer (CEO) Report

Kerry Sawyer gave the CEOS Executive Officer report on several topics related to CEOS. She began with the CEOS Montreal Statement, an expression of cooperation that is the backbone of the CEOS 2014-16 Work Plan. This work plan includes nine thematic areas and WGISS has key contributions for all of these.

* CEOS 2014 Priorities and Achievements: The CEOS 2014-2016 Work Plan was released in June with 80 objectives/deliverables that provide a sound and shared basis for measuring progress in a multiannual perspective. The final output of the CEOS Self Study (CSS) is a set of strategic documents that will guide CEOS in the future. It includes the CEOS Terms of Reference (ToR), CEOS Strategic Guidance (SG) (~10-year), and CEOS Governance and Process (GP) (~5-year), and ToR Elements for the CEOS Chair, SEC, SIT Chair, CEO and SEO. The CEOS governance will oversee adherence to agreed criteria for reviewing activities, in alignment with CEOS strategic goals, with benefit to internal and/or external stakeholders, and that are feasible and affordable. Kerry emphasized that WGISS should continue to support the development and operationalization of the GEOSS Common Infrastructure (GCI) and its CEOS-related elements. The Work Plan objectives and deliverables have been translated into a spreadsheet. WGISS is expected to populate the spreadsheet, giving a status report on the WGISS part.
* Timeline for Annual Update to the Three-Year Work Plan: Mid-November is the milestone to begin gathering inputs from working group chairs, virtual constellation (VC) co-leads, and ad hoc team leads to update and identify key objectives and activities for 2015-2017 with all external stakeholders, and participate in the GEO-XI Plenary meeting where the CEO will capture GEO’s priorities for the coming three years. In early December the CEOS-GEO coordination meeting will occur where the CEOS and GEO priorities will be discussed and synergies where CEOS can support GEO will be identified. After the coordination meeting, in early January, a draft Work Plan will be prepared to reflect inputs from CEOS entities and outcomes from the November GEO Plenary and the December CEOS-GEO Coordination Meeting. The draft Work Plan will be shared with the working groups (WGs), VCs, ad hoc teams, and CEOS secretariat to receive comments and inputs. In early February CEOS SEC comments will be incorporated and the Work Plan will be shared with the CEOS principals for virtual endorsement. The final version of the CEOS 2015-2017 Work Plan will be published in mid-February.
* Outcomes of SIT Technical Workshop: Kerry listed the actions from this meeting, highlighting in magenta those that are relevant to WGISS. WGCapD has an action to approach the WGs and VCs to identify PoCs for improved collaboration. All entities within CEOS are to develop two minute video describing activities and roles.
* Preparation for 28th CEOS Plenary: Kerry listed the expected outcomes for the 28th Plenary, and noted that a Tromsø statement will be developed by EUMETSAT and endorsed by the Plenary. The draft agenda is being reviewed, and Andy noted that WGISS is hoping to have a side meeting on emerging technologies.
* Preparation for GEO-XI Plenary: Kerry listed the details of the meeting.

## Systems Engineering Office (SEO) Report

Brian Killough, Systems Engineering Office, gave the following report of the activity of his office. He began by appreciating the assistance of WGISS in providing much of the detail for the Data Policy Portal. The portal has a few new features: CEOS interface (CWIC/FedEO), new missions, and updated client portal links. There are future plans to link with the MIM. Andy asked about definitions of the “CEOS interface”, and wondered if others had exposed an interface to their data. Brian replied that this is more of a promotion of the CWIC-FedEO interface, trying to get everyone to expose their data through one of these interfaces. Brian reported that CEOS operates 121 missions, with eight planned in 2014 and nine planned in 2015. Satoko commented that ALOS-3 will not be launched in 2015. CDTI and CONAE may provide a good opportunity for WGISS. Kerry suggested adding EC for Sentinel.

The COVE tool continues to evolve with new features, and added missions (total of 240 missions, 647 instruments). A new version of the Coverage Analyzer is due to be released in October. New overlays include Landsat WRS, Sentinel-2 tiles, and ASTER DEM. Current archive links include Landsat 7/8, SPOT 1-6, Pleaides-1A/1B, and Radarsat-2. Future archive links include TerraSAR-X, TanDEM-X, and RapidEye. Brian stated that linking the COVE tool to mission archives is extremely valuable to CEOS. GFOI and GEOGLAM would greatly benefit from using archive data to develop baseline maps and to determine future acquisition plans. Only the metadata for the images are desired. Only Landsat and Radarsat archive links are working well. Brian asked WGISS if it is possible to get archive links to ALOS-1/2, ResourceSat-1/2, Envisat, ERS and Sentinel-1A. Nitant Dube replied that as soon as the CWIC connector is up for Resourcesat-1 and 2 will be available, and ISRO has also done the connector for MOSDAC. Yonsook agreed to notify Brian when it is available, and added that providing the INPE ground station for Brazilian coverage may already be available. Satoko reported that ALOS-2 was just launched, so cannot speak to it yet. But for ALOS-1, if only metadata is needed, it may be possible to open the catalog server in its pre-operational phase. Mirko commented that there will be a presentation on Wednesday with more detail on ERS access through FedEO – the data is not quite available, as they are still cleaning and preparing (ERS). For Sentinel need they will need more time. Mirko will send Brian an email with the contact point.

Data Services Prototypes: Many developing nations lack the knowledge and infrastructure to access, process and utilize space-based data for local decision-making and national policies. Data Services Prototypes are NOT community portals, but rather, country-specific solutions to data storage and processing with secure access and ownership. Many GFOI countries would benefit from a dedicated data system to support national Measurement, Reporting and Verification (MRV) forest reporting. Similarly, such a system would also benefit GEOGLAM agriculture reporting. The SEO is leading several prototype data services projects to test ideas for the future.

The space data path chart: Acquisition Planning (COVE) 🡪 Satellite Acquisition🡪Data Discovery (COVE) 🡪Data storage and Data Processing🡪Product Generation🡪 National reports; where do countries fit in this path?

Data Services prototypes: Briand displayed a table showing pilot projects, participants, countries, datasets and processing tools, and whether they are for GFOI or GEOGLAM. The status of the prototypes is as follows:

* FAO – Contract between KSAT (Norway) and AMA (SEO contractor) in place (Aug 13) with a prototype delivery in late October.
* Kenya –Datasets provided to Kenya in June on a USB disk for in-country testing. SEO developing an online version for future testing and working with FAO on a demo.
* Colombia – SEO proposal for TanDEM-X datasets was accepted in June. Data has not yet been received due to issues at DLR.
* Asia-Rice – SEO has created an online system with 12+ Radarsat images and the INAHOR processing tool. JAXA is testing internally and will train Indonesia soon.
* JECAM – Project will start in late 2014 after requirements definition.

Brian shared his expectations for WGISS support:

* COVE links to mission archives: Resolve issues with CNES (SPOT and Pleiades) and DLR (TSX and TDX). Add new links to ISRO (ResourceSat-2), ESA (Envisat, ERS, Sentinels)
* Cloud Storage and Processing – SEO costs for Amazon storage are small ($20/month). Processing is more expensive ($0.35/hour). Supporting a few prototypes is reasonable (~$1200 each per year), but it is not possible for many countries and many years. Does WGISS have a recommendation for low cost, flexible and international cloud processing? Brian said he is putting together a package to go talk to Barbara Ryan in GEO – if GFOI and GEOGLAM is running well, they could pay the bill. But the SEO needs to provide options for implementation and recommendations.

## GEO Secretariat Report

Osamu Ochiai gave an update on GEO since WGISS-37, and also discussed the Data Management Principles Task Force.

The GEO Work Plan Symposium was focused on accelerating progress toward the 2015 Strategic Targets of the GEOSS, and preparing for the post-2015 era. The GEO Appathon was launched at Geospatial World Forum 2014; The Global APP development competition open to any non-commercial, 246 participants (104 individuals, 31 teams, and 49 countries). The GEOSS Asia Pacific Symposium welcomed the emerging initiative to integrate Earth observations in tackling the challenges of monitoring the complexities of the water sector in the Post-2015 Development Agenda with WHO, UN-HABITAT, UNEP, and GEO participation. Barbara Ryan was reappointed as GEO Secretariat Director.

Osamu listed status and highlights on GEO tasks IN-01, IN-02, IN-03, IN-04, and IN-05. Of key interest to WGISS is IN-03 (GEOSS Common Infrastructure) with 65 to 72 million discoverable resources (51 million for Data-CORE). The GCI requirements baseline was reviewed and consolidated. Satellite data assets are the dominant contributors to the GEOSS resources through the CEOS WGISS Integrated Catalog (CWIC). The CEOS International Directory Network (IDN) provided the source of registered data collections from Space Agencies in CEOS. Establishing interoperability between WMO Information System (WIS) and GCI is ongoing, as is the development of a set of “GEOSS Community Portals Guidelines” through some communities (e.g. GEO-BON, SAON). Also of interest to WGISS is IN-05 (GEOSS Design and Interoperability): Phase 7 of the Architecture Implementation Pilot (AIP-7) was launched in early 2014 to develop and deploy easy-to-use online (web and mobile) apps that demonstrate the value of standards-based access to EO data and services registered with GEOSS.

Osamu reported that the GEO-X Plenary agreed to setting up the Data Management Principles Task Force (DMP TF), which will support GEO in the definition and endorsement of common GEOSS Data Management Principles. The principles initially drafted by the GEO Infrastructure Implementation Board (IIB) will be reviewed and their feasibility assessed. Pending GEO-XI Plenary’s approval of the principles, DMP TF ToR may be extended and modified to include a mandate to develop implementation guidelines for these principles. Osamu listed the members of the task force, and explained that the foundational elements of maximizing Earth observations are Discoverability, Accessibility, Usability, Preservation, and Curation.

The Implementation Working Group (IPWG) is being developed in two phases: phase 1 a reflective, fresh perspective phase, and phase 2 for synthesis and formulation. The interim report provides an overview of stakeholders’ expectations, an analysis of the evolving context since the drafting of the first Implementation Plan, lessons from the past, and considerations for the future. The IPWG has proposed guiding principles for the development of the Implementation Plan for the next ten years. It recognizes that GEO will not change its overarching goal or vision, and sets out some objectives and then outlines main areas of activity foreseen for the coming decade. Through discussions during the first phase within the IPWG and through engagement with key stakeholders, some key considerations became evident which will help shape the development of the next Implementation Plan.

Osamu gave details for the GEO-XI Plenary, and showed an overview of the agenda.

Osamu requested the following contributions from WGISS:

* Strong cooperation with GCI (CWIC/IDN and GEOSS Portal and GEO/DAB)
* Input to IPWG (through agency, CEOS, or IIB)
* DMP TF (with cross cooperation with WGCV)

Richard noted that WGISS is ready to address the quality indicators whenever WGCV is ready. A regular meeting with GCI and GEO DAB is recommended in order to support the GCI. Yonsook reported that she had three teleconferences with GEO DAB, and Mattia Santoro knows how to get in contact with the CWIC team. WGISS has started an email list that will be used to report problems with any of the components that WGISS sponsors. The CWIC specific issues with GEO have a direct line of communication – there is an email list specifically for GEO. WGISS is also willing to set up regular teleconferences to address current challenges. WGISS participates in the GCI providers teleconference, and will ask to be put on the agenda. Osamu agreed that this would be a good approach.

Richard said this report was very interesting and in line with his own report. WGISS will go back with Osamu soon, to discuss concrete work in line with the GEO.

## GEO Infrastructure Implementation Board (IIB) Activities

Mirko Albani reported on the activities of the GEO Infrastructure Implementation Board (IIB). The objectives are to monitor progress towards achieving the 2015 GEOSS Strategic Targets and review “Infrastructure” tasks performance against the targets; to actively coordinate activities across “Infrastructure” tasks and establish cross-cutting links to the other Work Plan parts; and to advise on the implementation of “Infrastructure” tasks and provide guidance. Mirko listed the five infrastructure tasks (IN-01 through IN-05) already described by Richard and Osamu. It is through some of the components of the tasks that WGISS can contribute. Andy asked if IN-03 and IN-05 have CEOS participation. Mirko did not think so, but will check and request formal participation for WGISS in task IN-03 and IN-05. IN-05 is also where they are carrying out the AIP. These two tasks are very much in line with the work of WGISS. IN-05 is mainly working through the AIP. Through DSIG WGISS participates in IN-02. All the managerial work is managed by the OGC.

Yonsook said this is a good approach to test the ideas. But getting them implemented in operational systems is another step. What is different about what WGISS does is that the tests do become part of operational systems. For example OpenSearch uses the OGC specs, and are now implemented. CEOS OpenSearch best practices documents should be given to them.

Satoko suggested communicating with the lead of IN-05 about other activities to be included in addition to the AIP – WGISS can show its best practices documents, and also ask the GEO Secretariat how to publish the documents which are useful to many communities. Yonsook asked if the standards registry is still working. The value of the WGISS document is that it does provide a common way to access satellite data for significant numbers of datasets. Mirko suggested reviewing the inputs and outputs of IN-05 to see where WGISS can contribute.

CEOS and its member agencies are contributing to most “Infrastructure” tasks. Details on tasks activities and deliverables are available on the GEO Secretariat web site: <http://www.earthobservations.org/geoss_imp.php>. Dedicated presentations on specific “Infrastructure” tasks could be organized during next meetings with Task Leaders.

Additional IIB activities of interest for WGISS are the GEOSS Resources registration process and access, Community Portals Role in GEOSS, the GEO Label, the Data Management Principle Task Force (DMP-TF), and the contribution to the Implementation Plan Working Group (IPWG)

The ‘pledge’ of data or services to GEO requires the act of registration to provide basic information about the resource and its context (metadata). There is no need for registration if resource metadata is available in an agency, national, or professional catalog (i.e. INSPIRE, data.gov.uk) already registered with GEOSS or if the resource is already described in CEOS IDN.

Mirko described diagrammatically the data publishing in GEOSS. He also displayed the current GEOSS Infrastructure, and the proposed GEOSS workflow. The clearinghouse function will be embedded in the DAB, and the registration part will be simplified. Andy commented on the issue of duplication when a dataset is registered in the CEOS IDN and also in the GEODAB. It is time to identify the duplications and remove them. Andy suggested that WGISS publicize the data already available through the Federations so the DAB team knows; Mirko added that GEO is only one client of the federated data. As of May 2014, the GEOSS assets include more than 30 brokered data providers. Yonsook pointed out that these numbers are already out of date.

A Community Portal paper is being drafted with the purpose of assisting developers of Community Portals. As good and necessary as the GEOSS Portal is, communities still wish to develop their own interface to relevant data and services. Ideally Community Portals will utilize GCI for discovery and access of GEOSS resources. Mirko displayed a diagram of the Community Portal Role in GEOSS. Ken commented that the paper is out for review by the IIB and the AIP-7 participants but that a more general release is awaiting their input.  Continual updates are envisioned to capture on-going experience in using the GCI.

The GEO Label is a data documentation mechanism, a graphical representation of metadata completeness, and a representation of user experience. The different parts of the label were explained.

* How it is used: the GEO label can be created automatically by a web service. It uses a RESTful API that allows sending a producer metadata URL and a user feedback items collection URL
* How it interoperates with the GCI: the GEO label service is able to retrieve metadata directly from the Discovery & Access Broker and can be integrated in the GWP
* Proof of concept: demonstrated in the GeoViQua project and at GEO-X Ministerial Week, utilizing test implementations of the DAB and GWP
* Concerns about branding and about decisions on data documentation being taken out of the hands of data providers
* SIF recommends that the GEO label be made operational in the GCI
* Possible adoption and Implementation aspects under discussion at IIB

Yonsook asked if the label is in addition to the agency’s label. Mirko said perhaps the agency information is on the metadata; GEO cannot display the providers’ labels because they can come from so many different sources. It is a very delicate issue. Richard said this label can be very interesting to the RO. The need for each of the elements is exactly what is at the core of the system. Maybe there could be some exchanges to learn from this and to enrich the data in the RO.

IIB and IPGW:

* GEO Plenary (2014) has set up the Implementation Plan Working Group (IPWG) to “… undertake the preparation of a draft GEOSS Implementation Plan 2016-2025 for initial review at the GEO-XI Plenary and for acceptance at the GEO-XII Plenary …”
* IPWG will propose to Plenary for consideration a Strategic level document defining the scope of GEO's activity in three action areas – Advocate, Engage and Deliver.
* Implementation aspects (i.e. “How” including DMP-TF and DSTF) will be described in the supplement documents.
* IIB’s consolidated recommendations to IPWG were developed and delivered.

Current considerations at IIB:

* Clarify if GEOSS is a “System of Systems” or should evolve toward a less ambitious “Network of Systems”
* Definition of the target GEOSS Users and their needs is essential to make sure that the GEOSS Common Infrastructure can serve all of them (or redirect as needed)
* These aspects influence the entire architecture of GEOSS and drive the development/upgrade of common components (GCI)
* GCI operations sustained until 2015: need to assure the operational capability and necessary funding arrangements in order to continue and improve the GCI operations post 2015.

## WGISS Terms of Reference

Michelle Piepgrass revised the WGISS 5-Year Plan to create a document more in line with the CEOS Terms of Reference requirement; it should be at a high level and have specific goals identified. Everyone spent some time reviewing the document, and will provide input to Michelle for consolidation. Wyn led a discussion to uncover any major issues.

Suggestions made:

Add definition of CEOS.

Add key internal and external stakeholders.

Add CEOS organization how WGISS fits in it; also a generic structure diagram of WGISS.

Reduce the content.

Remove the text on work plan.

Harmonize and define the terms “access” “discovery”.

Remove definitive communication lines because those can change, and are covered in other documents.

 Only include long term objectives.

The content that paints the strategic picture should not be lost; it should be incorporated in another technical document.

Action WGISS-38-13a: Michelle, Ken, Jonathon and Wyn to finalize the “What is WGISS” text by end of October.

Action WGISS-38-13a: Michelle, Ken, Jonathon and Wyn to complete ToR for submission to CEOS Plenary.

## Future Meetings

Satoko Miura announced that WGISS-39 would be hosted by JAXA in Tsukuba, Japan, May 11-15, 2015. Tsukuba is 50 km from Tokyo and is the largest science technology accumulation site among the country (>300 research institutions, including JAXA). Satoko gave logistical information for travel from Tokyo to Tsukuba.

The venue will be the JAXA Tsukuba Space Center, on the first floor of the Headquarter Building. Wireless internet will be available in the meeting room. JAXA blocked 25 rooms at the Okura Frontier Hotel (<http://www.okura-tsukuba.co.jp/eng/>). Complimentary transportation between the hotel and the Tsukuba Space Center will be provided.

Registration is necessary (especially “in-person” participants). A JAXA member will support security procedures every morning, so please stay in the same hotel and come to the venue together by the complimentary bus. A no-host dinner will be planned.

Andy Mitchell reported that WGISS is looking for hosts for WGISS-40 and WGISS-41. A joint meeting with WGCV is not in the plan; all five days are needed, and a joint meeting would cut back on the available time. The linkages with the other working groups (WGClimate, WGDisasters, and WGCapD) are beneficial, so it was suggested using a joint venue instead of having a joint meeting. Jonathon Ross suggested that if the big data discussion continues, it would enable Australia to host a meeting. Richard added that hosting a meeting is a way to better understand the agency does. Yonsook added that topics of high interest like Big Data provide an opportunity for a workshop which brings in people from government and industry.

Mirko reminded WGISS of the Big Data conference at ESRIN November 12-14, 2014.

# Agency and Liaison Reports

## Academy of Opto-Electronics (AOE)

Chaoliang Wang gave the report for the Academy of Opto-Electronics, Chinese Academy of Science (CAS), describing the multi-temporal Landsat-MODIS images fusion via learning-based super-resolution reconstruction. He explained that temporal-spatial data fusion combines the spatial resolution of Landsat with the temporal frequency of MODIS. The proposed approach is based on change image super-resolution reconstruction via sparse representation. A change image is the difference between two observed images of the same area acquired at different times. He described the theoretical basis and the fusion algorithm. The low-resolution change image has the same sparse representation vector as the corresponding high-resolution change image. The algorithm includes dictionary training, Change Image Super-Resolution Construction, and optimal selection.

Chaoliang Wang presented experimental results and concluded that a new temporal-spatial data fusion approach was proposed based on learning-based super-resolution construction. Experimental results demonstrate that the proposed approach is more efficient for capturing spectral changes and spatial structure in areas of urban land with mixed pixels than STARFM. The performance improvement achieved comes with an additional computation cost in the dictionary training stage. Theyhope to develop method and criteria to transfer one type of EO data to another type when possible through research on remote sensing mechanism.

Chaoliang Wang also described Aerospace Application Coordination System for Emergency Response and Data Sharing (ArcSer) EO data sharing for the LuDian earthquake. ArcSer uses data from seven aerial and eight satellite remote sensing acquisition agencies, and many commercial spatial information service companies. Three emergency responses have been carried out since the system was established. Chaoliang described ArcSer’s response sequence to the earthquake. The forefront rescuer was invited to ArcSer’s consultation center, and took part in EO data acquisition discussion with disaster experts and spatial technical engineers. Thus the most suitable data acquisition task plans were formulated to meet the urgent need for disaster rescue and relief.

He concluded that it is very important to have an extensible and robust archiving and catalogue system to deal with multi-source EO data for emergency response. As UAV remote sensing system is flexible and easy to operate in a disaster area, it plays an important role in emergency response. It is necessary to strengthen communication with forefront rescuers, and to develop an individualized data acquisition schema that can be formulated to improve the efficiency and effectiveness of spatial resources used in disaster reduction.

Jonathon asked if the new technique for fusion has been published and peer reviewed. Chaoliang Wang replied that they have published it, and it has not been used in operation. Richard commented that ArcSer is very interesting, and may be a good experience report for the RO; it is completely in line with the goals of the WG Disasters. Richard will be requesting feedback on this.

## European Space Agency

Mirko Albani gave the European Space Agency (ESA) report. He began with a diagram describing the ESA EO programme. He described each of the Earth Explorer missions, which include GOCE, Swarm, SMOS, CryoSat, ADM-Aeolus, EarthCARE, and BIOOMASS. Copernicus is a European space flagship programme led by the European Union, and provides the necessary data for operational monitoring of the environment and for civil security. The Copernicus dedicated missions are Sentinel 1 to 6, and the long term operational perspective spans from 2011 to 2030. Sentinel-1A commissioning activities were completed on 23rd September and dissemination of all products to users will begin October 3. Mirko displayed early images that show oil spills, demonstration of land applications, flood monitoring, and earthquake mapping. Mirko described the free, full and open data policy adopted for the Copernicus programme, and gave the engineering view of data access.

Nitant requested that he make the slides available online. Jonathon thanked the European Community for this amazing program.

## Geoscience Australia

Jonathon Ross gave the Geoscience Australia (GA) report. He described the national geoscience agency’s size, and highlighted the priorities of building resource wealth, marine jurisdiction, community safety, and water resources through a network of remote sensing, geodetic and geophysical observatories. The agency’s responsibilities include maintaining and operating infrastructure, creating and assuring access to data, producing information that supports decisions, sharing knowledge and innovation, developing capability to harness geoscience, and collaborating nationally and internationally. The EO program applies EOS to governmental priorities. Jonathon described the EOS capability in terms of ground segment, expertise and data.

Jonathon described their participation in national leadership, and their international contribution, and listed their data holdings. GA is working on integration of these with CEOS. GA intends to put particular focus on sharing its experience in relation to the management and use of very large volumes of time-series EOS data.  GA has invested in development of a Data Cube focused on providing access to data and the ability to efficiently query and process it using High Performance Computing. GA is keen to share its work to date on the data cube, and will be continuing its development in areas of interest to CEOS members.

A few of GA’s priorities are economic analysis, preparation for data continuity, access to the Data Cube and technology sharing, national leadership, and international participation.

Richard asked if the Data Cubes could be useful for the RO. Jonathon replied that the difficulty is the size of the area; the data cube is targeted for big time series for an understood data, so it could be useful to the RO as they do want to cover for up to five years and see evolution over time.

## Global Spatial Data Infrastructure (GSDI) Association

Gábor Remetey-Fülöpp, Secretary-general, HUNAGI, presented an update on selected activities of the Global Spatial Data Infrastructure (GSDI) Association since WGISS-37, with some examples from EUROGI and HUNAGI.

GSDI received Special Consultative status at UN and described high-level cooperation with learned societies of the geospatial world. The group ad Disaster Recovery has new members, and is conducting an international symposium on disaster management with over 100 delegates. Gabor provided a list of GSDI links and publications.

Gabor listed GSDI Regional level member activities by EUROGI including NASA World Wind Europa Challenge, ESA/ROSA conference, EUROGI Position papers, imaGine, Big Geospatial Data Challenge, FOSS4G conference.

Gabor listed GSDI National-level member activities by HUNAGI, promoting EO apps competitions, activities with space industry, MaSat, Hungary and GEO news on developments, open source activities, FÖMI, ICA Working Group Workshop on Digital Approaches to Cartographic Heritage, Fény-Tér-Kép (Light-Space-Image) Conference, Innotrends 2014, Contribution to the Internet of Spaces Days, and the EURISY Workshop.

In conclusion, Gabor commented that interoperable spatial data infrastructures and related services are enabling tools for EO applications. The EO industry is transforming into a geoinformation industry. GSDI and IGS are open for the EO players to join the geospatial information community. GSDI regional and country level members are playing an active role not only supporting EO applications by their members for the societal benefit areas but also providing awareness raising by arranging international networking, regional projects, conference sessions, thematic workshops and promoting competition challenges from local to regional in EO in close cooperation with space agencies such as NASA or ESA.

## Japan Aerospace Exploration Agency (JAXA)

Shinichi Sekioka gave the Japan Aerospace Exploration Agency (JAXA) agency report. He began with the announcement of the successful launch of ALOS-2 May 24. First images have been released, and PALSAR-2 products will be released in late November. Shinichi listed the specifications of the Advanced Land Observation Satellite (ALOS-2), and also for the Phased Array L-band Synthetic Aperture Radar2 (PALSAR2). He displayed early images. Shinichi also announced the launch on February 28 of the Global Precipitation Measurement satellite, and noted that the GPM products are available from the G-Portal. Shinichi described the Global Satellite Mapping of Precipitation (GSMaP), a project to increase knowledge or tropical cyclones and typhoons.

Shinichi displayed a diagram of the current and planned JAXA missions for climate, water, and disaster.

Martin asked for if all GPM data is on the G-Portal; Shinichi replied that are distributed as the cal/val phase is over.

## National Aeronautics and Space Administration (NASA)

Andy Mitchell gave the National Aeronautics and Space Administration (NASA) agency report. He described NASA Earth Science, listing their six areas. He also explained the role of NASA’s Earth Observing System Data and Information System (EOSDIS), whose strategic plan is to advance Earth system science to meet the challenges of climate and environmental change. The aim is to present NASA’s EOSDIS as an interoperable system of systems where users can select, view, interact and download the data they need transparently from all subsystems in support of interdisciplinary Earth Science research. EOSDIS provides archive, processing, access, data management, user services, and also network and data transport responsibilities.

Central reusable capabilities are heavy on metadata services, user tools, metrics systems, and user registration systems. He announced the release of the new Earthdata Search Client which will showcase CMR, GIBS, OPeNDAP, and other services, and improved the user experience for Earth science data search, discovery and access. Andy explained that ESA’s DHuS Data Centre(s) plug into EOSDIS. For Sentinel 1, 3, 5P NASA will leverage proven mirroring and redistribution capabilities. Andy displayed a diagram of NASA airborne campaigns and Earth Science Data Operations and Earth Science missions (operating and planned through 2021).

## National Oceanic and Atmospheric Administration (NOAA)

Martin Yapur presented the National Oceanic and Atmospheric Administration agency report. He listed the four main priorities: Make communities more resilient, evolve the weather service, invest in observational infrastructure, and achieve organizational excellence. He also listed recent accomplishments: GOES-R and JPSS, Jason-3, DISCOVR, COSMC-2/GNSS RO all on track.

NOAA continues to build its value tree, with more than 100 data sources, classified into various mission areas. They seek to discover the user value of each of these, and study system impact versus cost. He showed a JASON product example to identify the products and impact levels associated with the JASON satellite.

Martin noted that NOAA data are unique, valuable, and irreplaceable, with many observing systems and a scope from the bottom of the ocean to the surface of the sun. NOAA also has multiple external engagements (CEOS, NSF, ESIP, OGC, USGEO, GEO, and DATA.GOV). He listed several data accessibility, usability, and data preservation activities, adding that NOAA is participating in the inter-agency activity BEDI. NOAA is also working to improve weather forecast, and is releasing a new weather forecast model.

Martin confirmed that NOAA data is open, free, and downloadable, and will be available to EUMETSAT, and other broadcast services. Jonathon mentioned that coordination is required to get it to other countries.

# WGISS Projects

## CEOS Water Portal Project

Satoko Miura (JAXA) opened the session with a remark that there will be new updates to the project by the end of this ﬁscal year. The Water Portal continues its operation at <http://waterportal.ceos.org>.

Shinichi Sekioka described the CEOS Water Portal as a distributed data system component of Data Integration Analysis System (DIAS) to provide easy access to a variety of hydrological data, connecting existing components such as data centers, scientists, and general users. He described the type of data and services available, and displayed a goals diagram and a partners’ diagram, and listed available data.

Shinichi announced two new data partners: NOAA/NCDC Climate data, and GEO DAB (Data Access Broker) connecting with GEOWOW via GEO DAB, and includes the important River Discharge data. A diagram of the new architecture development was displayed. OpenSearch two step search will be implemented using internal servers (based on "GI-Cat"), instead of using IDN (for the first step) as proposed at the last meeting. Future plans are being discussed, and include development of synergy with other components of DIAS, and a SSO enablement within DIAS (Water Portal is a component).

Wyn asked if there are any actual users yet. Satoko replied that about 200 users per day are visiting the website, but they are not sure if they are accessing data. The portal is still in development and is part of a broader system at Tokyo University. They need to connect to operational bodies, and are still seeking the way to an operational phase. Nitant said that he can provide capability for forgotten passwords that is part of the sign on page. Andy asked what kind of SSO technology they are using, noting that coordination with others is needed for the decision to be made. JAXA plans to achieve SSO between the water portal and other DIAS components this year. Nitant asked if real time and time-series capabilities can be added to the portal. Satoko said they are already providing these, and Yoshiyuki added that there is real time visualization as well as simple download.

## Disasters Recovery Observatory (RO) Project

### Introduction

Richard Moreno gave an introduction and description of the Recovery Observatory project. He explained that version 1.0 will be very limited and in time for the CEOS Plenary (end of October). Version 2.0 is expected for the end of July 2015. The Recovery Observatory Infrastructure organization includes the WGISS participants (NASA, NOAAA, ESA, CNES, UKSA, and China/Codata), WGDisasters, experience from KalHaiti, and a CNES contractor (AKKA) for development and integration.

Richard listed the events spanning the project inception (Q2 2014) to present time. He explained that governance, hosting, exploitation, and maintenance are out of scope for this project.

### RO Presentation to WGDisasters

Richard presented highlights from the presentation given by Steven Hosford, chair of RO Oversight team, September 11, 2014 to the CEOS Working Group Disasters (WGDisasters).

Space agencies already organize the response part of the cycle (International Charter, Sentinel-Asia) and follow an ad-hoc approach on scientific requirements for a specific zone or disaster/theme. After the initial response, space agencies have little or no coordination for the post-crisis part of the disaster management cycle. The RO proposes an infrastructure providing access to EO data of use in the Recovery phase (non-profit activities). Data should include the data acquired during the response and a regular (every several months) acquisition during Recovery. Links to local users and to international organizations.

The Recovery Observatory, spanning three to five years, would organize EO data from response phase and pre-disaster in repository, and plan coordinated acquisitions to support built area damage assessment, natural resource and environment assessment, reconstruction planning, reconstruction monitoring, and change monitoring.

The implementation proposal from the Oversight Team created summer 2013 (CNES, ESA, JAXA, NASA, ASI) is one Observatory as part of Observation Strategy 2014-16, building on the success of Charter, Sentinel-Asia, and KalHaiti project. The detailed analysis is complete and was approved by the SIT in Toulouse, April 2014.

Next steps are to determine conditions for triggering, infrastructure establishment, and liaison with DRR stakeholders, generic planning, cold storage, triggering, operations (3-5 years), and closing. The first Recovery Observatory ready for triggering by 28th Plenary (October 2014), with reparation phase completed July 2015.

The focus of the current effort is outreach to stakeholders and involvement in partnership, with key DRR partners to support triggering process by recommending appropriate event and coordinating relationship to national end users. For EO data licensing the focus is to move from an ad hoc diverse licensing approach to standardized license. The proposal is a new “open” license tailored to image data which would be adopted by data providers, and to considerably simplify the requirements on the IT infrastructure (management of one license/set of access constraints). The focus on IT infrastructure and related planning is working through WGISS, CEOS to propose solid IT approach to organize access to data and support usage.

The Recovery Observatory Oversight Team has made informal approach to World Bank, Red Cross, UNISDR and UNOSAT in spring 2014, and they expressed interest in pursuing discussions. A second approach to stakeholders was made formally in summer 2014: UNISDR, UNOSAT, World Bank/GFDRR, UNDP, Red Cross and UNEP were invited to a consultation meeting to be co-hosted by World Bank in Geneva, either 6 or 7 October, to be held jointly with a meeting of Recovery Observatory Oversight Team. The first in-depth discussion occurred in early September with World Bank/GFDRR who showed a strong expression of interest and exploratory discussions on development of institutional relationship in context of Recovery Observatory. Next steps: face-to-face meeting in Geneva, formal path forward for Observatory triggering, though it is possible that partner commitment will only follow late 2014.

The proposed triggering process is as follows:

1. Charter activation - Oversight Team members or international DRM stakeholder/partner consider whether the event is catastrophic and warrants Observatory.
2. If event is considered to be a candidate event in the subjective judgment of any Team member or Associate Member (International Stakeholder), One Oversight Team member or Associate Member calls teleconference to propose triggering.
3. Oversight Team examines predetermined scenarios developed during preparation phase and looks at triggering criteria, which would include considering advice from DRM stakeholders (and commitment in kind or resources for value added support), looking at the scale and scope of the disaster, and looking at benefit to be derived from establishment of Observatory.
4. At least three Oversight Team members and at least one Associate member, with the agreement of the CEOS Executive Officer in consultation with the CEOS Chair and SIT Chair, representing critical mass/varied contributions of data (according to what was deemed to be required in the preparation scenario), decide they would like to trigger the creation of the Observatory.
5. The Observatory is announced within CEOS with a call for participation.
6. The Recovery Observatory Oversight Team generates a situation report and acquisition strategy to support triggering.
7. The Observatory is announced publicly, jointly with a DRM partner.

The core user community for the Observatory is at the national and local level. In the first weeks following a disaster, the international DRM stakeholders will establish a liaison with the National End User and will support the generation of imaging requirements according to a product list established in coordination with the national end user. After the initial response phase is completed, the Observatory will work directly with the national and local end users, ensuring requirements meet expectations and that products respond to needs; the international DRM stakeholders will support and facilitate this process. Once the Observatory is established, the Oversight team will work with DRM stakeholders and national and local authorities to develop a legacy strategy, ensuring data and products are useable for long-term recovery and reconstruction efforts, and support resilient reconstruction.

The proposal of a specific data license is based on the objective to facilitate the distribution of (satellite) data by defining a data license based on elements from current licensing arrangements used by commercial providers. License could be based on open source examples such as GPL (GNU Public License) or Creative Commons. License could be used by agencies when provisioning data for RO/Disaster pilots.

Valery asked for elaboration on the difference between the RO and the WGDisasters. The WG encompass many aspects of disaster, and RO is a specific project focused on recovery; the scope of the working group is much wider, though when the RO is triggered, it will likely be their main activity. Andy asked if the RO is an architecture that will be used by the WGDisasters as a project that they will be able to use to give homogeneity to the Disaster activities. Specific projects that pertain to other working groups will remain within those groups.

### Recovery Observatory Version 1.0

Mathieu Gond gave a detailed presentation on the Recovery Observatory. The reference projection is predefined as EPSG: 3857. The RO Global Interest Zone (GIZ) is defined once and for all, with access to available OpenSearch EO catalogs, and other like SPOT-IMAGE / Digital globes. RO data hosting is limited to a subset (user data) and is focused on a subset of RO cartographic search/viewer features. RO collaborative features will be included in the next versions.

Due to time constraints only a subset of the expected features will be available. These include cartographic based data query, limited user contribution, upload, and minimalist user management (administrator/user). A use-case driven approach will be used for the development of this first version: scenario selection, well known software reuse, and development of the missing part in order to provide an operational version for CEOS Plenary. Parallel activities are formation of libraries and software reuse studies, RO specification finalization with CEOS partners, and end-user scenario finalization with CEOS partners. Mathieu described the following use cases:

Use Case #1 – RO initial context setup: triggered by CEOS decision to set up the RO for a given disaster. The RO instance will be installed by an administrator. The administrator will setup RO initial context: Define the Global Interest Zone, set range dates, choose disaster type, select EO and other data sources, and import external data in a given layer. Depending on disaster type, data will be ordered from local and global providers, specifying frequency and first occurrence. The administrator will be able to select other sources which will be rendered as layers on the cartographic web application. Static referential sources (toponyms, roads, demographics) will also be requested. A geographical footprint will be automatically tagged against location, land cover, etc. using iTag (<https://github.com/jirom/itag>). iTag (Opensource and extensible) can tag a footprint with information on continents, countries, cities, regions, states, geophysical plates, volcanoes, land cover, population count.

Use case #2: Data query interface. After RO initial context setup is done by the administrator, the users will access data through a web application based on the map.

Use case #3: User contribution upload. Users can upload files on the server using HTML (limited to a few formats); user will send a mail to the operator for ingestion. The operator will trigger ingestion supported with product metadata extraction and will fill missing metadata. For each product to be ingested in the system an associated metadata file must be provided in GeoJSON format (<http://geojson.org/>). The data model is based on the EO OpenSearch data model. RO will provide a toolkit (Python) to support data providers (or RO administrators) in extracting metadata from well-known product types (DIMAP, Shapefile, and geoTIFF).

Use case #4: EO catalog harvesting to host the metadata copy inside the RO which will be enhanced automatically and by RO users. Metadata catalog harvesting will be implemented as a scheduled service with defined parameters (GIZ). The RO data model starting point is reuse of THEIA data model, CEOS OpenSearch Best Practice document, and data annotation with iTag.

Mathieu gave a demonstration of the minimal version of the RO. He commented that a lot of software needs to be plugged together, and make them work together.

### Discussion

Chaoling Wang commented that his agency has something similar and that several issues have been raised. It can be difficult to determine the end of a disaster recovery observatory; how will the RO determine the end date? Jonathon said that this would be determined by WGDisasters. The quality of the EO data is important to disaster data sharing, including percent cloud cover, signal-noise ratio, resolution, etc. How does RO control the data quality? This is also in the scope of Disasters Charter, not WGISS, and WGDisasters will provide the management of the RO. When a space agency uploads the data for disaster recovery, FTP may be another choice for the system to operation. It may be useful for the end user to provide data type list querying function when considering search service. As the system archives multi-resource EO data and there are many kinds of metadata, XSLT (eXtensible Stylesheet Language Transformations) technology may be helpful for extracting metadata from metadata files.

Valery asked how the project interacts with the existing disaster recovery projects (UNSPYDER, Sentinel-Asia); he would like to see how it fits within the overall structure, and how it interacts with other activities. Mathieu said that so far they have only had input from CNES, and that the team is seeking input from WGISS to enrich the scope of the project. Andy commented that the RO is to continue after the Charter is closed, providing a longer term provision of data.

Andy initiated a discussion on data latency, asking when data is considered near-real-time, and when it is science data. At NASA near real-time (NRT) is three hours to seven days. Would the RO want NRT, science data, or both? Mathieu suggested both; though the reality is that after seven days the NRT is no longer available. One idea is to follow the Facebook model where everybody puts in data, so there is a community of data. Andy also commented that in the USA the requirement is to export data in the JSON format, and wondered if the GeoJSON is a best practice. Mathieu replied that GeoJSON is nice because it has the geometry and can be extended; it has only two or three keywords. The tool RESTo is extendable so there can be many connectors.

Wyn asked if there will be more than one RO, and if so, will hosting agencies take ownership and maintenance of them. CNES is hosting an example, and agencies will need to develop their own systems.

Nitant commented that right now geoTIFF is supported, but in future they may support other types of data and agencies need to know how to extract the metadata. Nitant asked if there is any interface; currently the data access is by HTTP. Jonathon said this is approved for one instance, and others will provide value added.

Yonsook asked if all the data will be ingested, and what is the order of magnitude of the scale of data. Mathieu said that the RO is meant to be a lightweight system, so they will be ingesting the metadata, and that for KalHaiti they had 1000 observations, and expect 10 times the amount for the RO.

Valery raised three issues: Where should observatory be implemented within CEOS groups (out of WGISS, WGDisasters). Andy replied that WGDisasters will manage it, and WGISS is only involved in the data architecture. Valery asked how the interaction will be managed in terms of governance and procedures. Andy replied that this question is at the level for the SIT and Plenary, not WGISS. His third question was related to optimizing national approaches. For example, Russia has something similar to the RO, and the team should look at the links among those. Andy added that there is a need for coordination of national assets, to avoid duplication of effort.

Nitant suggested that rather than ingesting the data, they should be setting up the access, and WGISS should work on a direct access to the data. The ingest model for the actual agencies would be a huge burden, and also latency issues would be a problem. Yonsook noted that WGISS is building the backbone of this for search and access, providing the tools to access the backbone in such a way that the RO portal would “look” like they have the data already. Michael Morahan commented that many of the assets are already searchable. There was additional discussion of the pitfalls of harvesting; Mathieu will prepare an email with the project team’s response to this topic.

Andy commented that there are possible linkages between groups at two levels – technical level and strategic level. Additional comments were that the RO should provide capability for end user to add layers, and it would be useful if WGISS members could compile a list of global and commercial data that can be exposed in RO.

**Action WGISS-38-1**: Richard to discuss with WGDisasters how the RO interacts with other disaster systems, not just at the agency level, but also UNSPIDER, Sentinel-Asia, Google (this is a strategic level question).

# Interest Groups

## Technology Exploration Interest Group

Andy Mitchell and Satoko Miura introduced the Technology Exploration Interest Group (TEIG) session.

### CNES - Big Data Technologies (Hadoop)

Benoit Frezouls gave a presentation on the CNES Hadoop system for Gaia data processing. The Gaia mission is an ESA astrophysics mission with the objective to build a map of the sky, and must manage billions of objects, so normal techniques are not able to handle the volume of data. There are six receiving stations in Europe, and Benoit described the Gaia/DPAC organization. Technical challenges include processing power, volume of data to be processed. Benoit detailed specifics on the data processing, where they use a Main Data Base (MDB) concept, which he described, and also the Façade concept.

CNES set out to identify a scalable solution able to cope with hundred/thousands of concurrent clients, and handle Petabytes of data/billions of rows. The solution must be affordable, able to handle complex queries, build the scientific chains input data, and be integrated into the PCC framework with maturity and support. A number of solutions were explored, and a relative performance comparison of the different solutions on the same dataset and the same test cases was conducted. The solution chosen by the DPCC was Hadoop, with a batch execution framework and an HDFS parallel file system. It has the advantages of speed, integration, and scalability.

Cascading simplifies the use of Map/Reduce paradigm (abstraction layer) with simple query API (directly in Java). Hadoop jobs (data processing) = flows in Cascading converted ‘on-the-fly’. There is only a small performance overhead of five percent.

Benoit gave an infrastructure overview and described the operational and validation hardware.

Fine Hadoop tuning is needed and is complex. He discussed high memory consumers Vs high CPU consumers, optimizing the calculus nodes usage, and ensuring prioritization between the different processing chains (queues).

Richard thanked him for his presentation and encouraged the participants to email him with any questions.

### Geoscience Australia Data Cube

Jonathon Ross introduced his presentation wondering if big data is a problem or an opportunity. The massive growth in both the variety and volume of direct acquisition and transmitted spatial data presents a number of challenges to researchers who want to realize the potential of the information contained within this resource. Geoscience Australia itself maintains a deep archive of Earth Observation data accumulated over the last 35 years. During this time, they have accumulated over 200TB of downlink data from their ground station network. Considering each of the pixels represented by those acquisitions as observations, it amounts to over 100 petapixels of information.

The Copernicus program is amazingly visionary, and the next generation of geostationary satellites represent an amazing opportunity and challenge. MODIS imagery is acquired every 10 minutes; in one sense, it looks like data is a ‘problem’ to be dealt with. On the other side of this deluge of spatial content is a growing expectation by decision-makers for on-demand access to data from a diversity of sources for use with other content without a reliance on time-consuming and costly integration processes. The expectation amongst decision makers is for geospatial data integration on-demand. To enable this, a common analytical platform is required that can link very large multi-resolution and multi-domain datasets together to enable application of the required analytic processes.

Traditional approaches to acquiring, managing, distributing and even using this data have made it challenging to realize the true value of this data. Traditional models where acquired data is only prepared to the point of being useful when somebody asks for it, create a significant barrier to entry for those businesses, agencies or researchers thinking of a new initiative. And even when the interest is there, the sheer amount of time and cost involved in the preparation process can be prohibitive. The use of approaches such as mosaicing, where the raw data is ‘merged’ from multiple passes of the satellite to produce intermediary products that are manageable and usable on commodity ICT infrastructure is an achievement. But an unavoidable effect is that often valuable data is ignored; for example, reducing a potential 45 observations of a spot on the ground from the Landsat series down to just 1 that has been judged the ‘best’ for that year. The storage of much data on tape archives, whilst ensuring data is preserved, simply does not work in a world where information is needed ‘anywhere’ and ‘now’. Such approaches often result in the product of a product that provides a ‘point in time’ perspective, with significant challenges to creating the sorts of dynamic information now essential to businesses and governments.

The science objectives are to engage with dynamic data: growing in time; subject to modification (existing data) and insertion (new data), and for precise observation time for each discrete observation. Science users also need every observation and avoid mosaics - full time series are key for many applications, and they need to reliably compare over long periods of time for pattern analysis. Users need to go ‘beyond’ the pixel, look sideways, mashup ‘scales’ and ‘types’, work iteratively at continental scale, use numbers not pictures, implement science principles. They must use statistical analysis, high performance computing, and scientific algorithms to ‘see’ things that the naked eye cannot see, in order to gain insights that are simply impossible by eyeballing images with the ‘naked eye’.

The challenge of big spatial data has led to a response in the form of development of the Australian Geoscience Data Cube. The AGDC (data cube) is a common analytical framework composed of a series of data structures and tools which facilitate the organisation and analysis of large data collections. A common framework is required that will link very large multi-resolution and multi-domain datasets together and to enable the next generation of analytic processes to be applied.

Building Block 1: The Australian Geoscience Cube is a collaboration between NCI and CSIRO. The Australian Government recognized the need to invest in High Performance Computing research infrastructure. As a result Geoscience Australia has access to the country’s largest supercomputer. It is only with access to this sort of computing power and storage that the work AGDC is performing is possible. It is accessible to their key partners, CSIRO and the Bureau of Meteorology, meaning it is not just a big computer, but a place to do big collaborative science.

Building Block 2: Solid science - taking the data to comparable quality assured measurements; physics based.

Building Block 3: Shifting to processing on receipt and managing towards a defined ‘collection’.

Building Block 4: Remote sensing data is typically both spatially and temporally irregular. Data is “clumped” both spatially and temporally and, hence, not suited to the monolithic array approaches traditionally employed due to large volume of no-data pixels. The AGDC arranges the data spatially and temporally to allow efficient large-scale analysis. “Dice’n’Stack” method is used to subdivide the data into spatially-regular, time-stamped, band-aggregated tiles which can be traversed as a dense temporal stack.

Building Block 5: The quality of each observation is represented by a set of bitwise tests, including the contiguity of pixels, and information which enables a user to determine if the observation is likely to be clear, not shaded or obscured by clouds. This approach took a lot of its inspiration from the MODIS approach to quality representation.

Currently the AGDC includes surface reflectance data for Landsat and MODIS, Landsat fractional cover, several versions of the 1” SRTM DEM, ASTER mineral products and the Australian Geographic reference image, with AVHRR and MERIS to be included shortly.

The Data Cube paradigm has enabled GA to do not only undertake this analysis at a regional scale for the first time, but also to scale this to produce a comprehensive national product. They can now analyze 15 years of data, across the country, representing every 25 square meters in under one day. They can produce it in an ongoing basis, enabling the production of updated information as new data becomes available, and enabling researchers to interact with the data iteratively to improve quality and try new ideas. In future, industry and others can build value-adding products on top of this data by leveraging the services-based approach adopted.

Future challenges include collection management and going from continental scale to global scale. These are areas where GA can contribute to WGISS.

### Big Data and Cloud Computing for CEOS

Brian Killough, CEOS SEO, gave a presentation on the CEOS needs for Big Data.

In terms of the Australian data cube concept, the CEOS Systems Engineering Office (SEO) proposes to work with Geosciences Australia on a “data cube” concept for enhanced use of space datasets. Data Cubes allow users to query data in time and space without the need to download individual scenes. This concept has proven very effective in Australia. Landsat spectral band data, cloud information and other ancillary data are stored in geolocated stacks to allow easy time-series analyses and classification studies. A supercomputer would create the data cube with historical data from Landsat (-5, -7, and -8) and MODIS. An Advanced Programming Interface (API) would allow users to conduct queries on the data cube and generate products, such as forest cover report (change detection), cloud-free crop maps, etc.

The SEO and GA have decided to focus on Tanzania for a Data Cube project. Tanzania has both GFOI (forest cover) and GEOGLAM (agriculture) interest. GA has provided SEO with a “Virtual Data Cube” to install on a PC for testing. Preliminary plans are to use only Landsat data and expand to MODIS and Sentinel later. Sample products would be NDVI, Fractional Cover and Anomaly detection. A meeting will be held at GA in early 2015 to work out additional details and train the SEO to create a data cube and implement API tools.

The National Computational Infrastructure at Australian National University houses is the world’s 32nd largest supercomputer. Tanzania scales to 13% of Australia 26TB and 47 Landsat scenes. The SEO is uncertain where it will create the Tanzania Data Cube. If the Data Cube approach is promising, then we will need to consider regional “Big Data” Nodes to create and host other country-level data cubes.

Does WGISS have any suggestions for Supercomputer resources to create the Tanzania Data Cube? What are the WGISS thoughts about “Big Data” Nodes for creating and hosting other Data Cubes? Richard replied that WGISS does not know of any available supercomputers. Further investigation is needed into the concept of Data Cubes.

Brian concluded that the first test the concept is for small Tanzania case; when the Sentinel data become available the scope can expand, so discussions of scaling up will be needed.

### Big Earth Data Initiative (BEDI)

Andy Mitchell gave an update presentation on the NASA BEDI, an effort that aims to make EO data collections more available and useful to users using open standards and protocols. The BEDI design approach is:

* Discovery: Make products known to users, and find and document existing products in Echo/GCMD Started with the request for information from ESDIS. Use the Global Imagery Browse System (GIBS) as a means of discovery.
* Access: Make products readily available to users, and make popular products available – preferably in near real time. Implement improved cataloging of products. Provide improved distribution/management tools
* Usability: Provide the tools to visualize the products, and provide easily used tools or mechanisms to access and aggregate data. Include enough supporting information that the consumer knows what the information means.

Andy described the activities needed to meet the design approach.

For each of the 12 Societal Benefit Area (SBA) data sets that were identified, they will provide current metadata to ECHO and GCMD, assign and register DOIs, make data available online via OPeNDAP or some other service based on open standard protocols like OGC, and make imagery available in GIBS. Andy listed the 12 Societal Benefit Areas that are part of the BEDI Initiative, and gave statistics on the size of ECHO and the Common Metadata Repository (CMR).

Richard asked if OPeNDAP will we used only for gridded data; Andy replied that they are hoping to use it for everything. Michael Schick commented that it is for subsetting. Nitant said OPeNDAP provides the capabilities; how do you fine tune OPeNDAP to handle big data. Andy said this is something that they are analyzing. Satoko asked if BEDI is multi-agency and is there funding. Andy replied that yes, there is some funding, and it is multiagency.

OPeNDAP could be very useful topic for Technology Exploration discussion.

**Action WGISS-38-12**: Andy Mitchell to write a summary of what NASA is doing with OPeNDAP (UNIDATA) and to organize a session on OPeNDAP at WGISS-39 (inviting presenters from OPeNDAP).

### TEIG White Paper

Andy and Satoko reported that they composed a white paper on the Technology Exploration Interest Group session from WGISS-37, and put it on the WGISS website. They have decided to have these summaries once per WGISS meeting.

The question was raised as to how detailed these papers should be. Jonathon commented that these are common issues, and since not everyone will implement them in the same way, a valuable addition could be a list of “principles” to follow, and a list of definitions. He added that his impression is that WGISS has made enormous progress in catalog integration, and a white paper drawn on lessons learned would be very useful – a short document to bring up the visibility as a focus, and to effectively support GEO CEOS needs by thinking in these paradigms – to get the data used more widely.

### CEOS Plenary Side Meeting

Andy noted that WGISS is organizing a side meeting on technology at the CEOS plenary. The attendees will not be very high level people, so the discussion should explain from a scientist point of view what big data is, what the challenges are, and if overcome, what the benefits and results would be. The interest group will develop slides and a handout that includes a description of WGISS (taken from the WGISS home page). Part of the purpose of the side meeting is to focus on the CEOS objectives for data – and how WGISS helps to identify and fill gaps in the supply chain and to facilitate its discovery and access.

JAXA is also hosting a side meeting on data applications; this will be a good follow to the WGISS side meeting. The WGISS Exec will work together to optimize the scheduling and handle the logistics of the meeting.

It was suggested that the “WGISS” header is missing from the WGISS landing page, and that “exploitation of data” should be a key phrase on the WGISS description. The Technology Exploration IG is trying to explain in the Plenary what WGISS does – the side meeting is a good place to start explaining. Jonathon agreed to look at the text of the page and suggest revisions. Nitant suggested that the next logical step is how everyone can access the data, and how to process the data as desired.

**Action WGISS-38-14a**: Andy and Jonathon to investigate an existing project for a few GFOI or GEOGLAM prototypes that would benefit from data exploitation that WGISS could promote.

**Action WGISS-38-14b**: Technology Exploration IG to work with Brian to investigate options for supporting cloud processing, so Brian can present these to Barbara Ryan with a common voice between WGISS and SEO.

## Data Stewardship Interest Group

Mirko Albani introduced the Data Stewardship Interest Group (DSIG), and listed the presentations for the session. He also described the group goals, scope, and membership.

### DSIG Plan of Activities

Mirko presented the current DSIG activities:

1. Share information and experiences during WGISS meetings through reporting from CEOS agencies about their investigations, developments, studies, and lessons-learned relating to EO data management and stewardship, having dedicated sessions on specific topics.
2. Draft common cross-agency best practices or guidelines on data management and stewardship for possible adoption by WGISS. Preliminary list includes Preservation Workflow, Data Preservation Guidelines, Preserved Data Set Content, E2E Level-0 data consolidation, Purge Alert, Persistent Identifiers, Appraisal, Archived data media transcription.
3. Conduct joint activities and/or pilot projects on specific data stewardship topics.
4. Contribute to GEO and standardization activities in the domain of data management and stewardship in the frame of Component C1 - Advances in Life-cycle Data Management within Task IN-02 Earth Data Sets. Propose WGISS best-practices/guidelines drafted by DSIG for adoption in GEO and possibly as input to standardization bodies (e.g. OGC/CCSDS).

Mirko displayed a diagram of the data stewardship best practices and guidelines cycle and a proposed timeline.

Mirko reported that the GEO Data Management Principles Task Force (DMP TF) drafted a set of high level common GEOSS Data Management Principles covering the entire data life cycle from planning, to acquisition, quality assurance, documentation, access, archiving and preservation. Mirko Albani and Richard Moreno are members of the DMP TF. There are two new actions:

1. Review IN-02-C1 (Advances in Life-cycle Data Management) Task Sheet on GEO Secretariat Web site after GEO XI Plenary to align participation, activities and outputs with DSIG.
2. MA/RM to inform DSIG members about future DMP TF discussions/outcomes.

Mirko also reported that the WGISS-37 actions for the DSIG are closed.

### DSIG Data Stewardship Best Practices/Guidelines

Mirko Albani presented DSIG data stewardship best practices and guidelines, noting that they are working on four core documents:

1. Preservation Workflow
2. Data Preservation Guidelines
3. Preserved Data Set Content
4. E2E Level-0 Data Consolidation

Mirko gave specific definitions of data records, associated knowledge, representation information, E2E consolidation, preservation, curation, and EO data stewardship. Andy commented that these are terms that WGISS should promote in a glossary of terms.

**Action WGISS-38-11**: WGISS to develop a glossary of terms.

Mirko described the preservation workflow, with its inputs and outputs. The Data Preservation Guidelines Issue 2.0 is available at <http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_Issue2.0.pdf>. Comments are being collected and a consolidated draft will be circulated to DSIG for review and then to WGISS chair for approval cycle.

The Preserved Data Set Content (PSDC) Issue 4.0 is available at: <http://earth.esa.int/gscb/ltdp/LTDP_PDSC_4.0.pdf>. It provides a description of the composition of the EO “Preserved Data Set” indicating whatto preserve in terms of data and associated knowledge during all phases of an Earth Observation mission. A consolidated draft will be circulated to DSIG for review and then to WGISS chair for approval cycle. Mirko listed four principles.

E2E level-0 data consolidation: this process defines a set of actions to be applied to generate a single, consistent, consolidated and validated Level-0 “Data Record”. Document being drafted by ESA and LTDP WG will be circulated to DSIG for review and then to WGISS chair for approval cycle.

Mirko gave a timeline for data stewardship best practices/guidelines. Final presentation and formal approval is expected May 2015, with the intervening period for review, feedback and revision.

Wyn suggested a document tree since so many documents are feeding into others. This will help understanding the relationship between. Satoko added that since these four documents may be finalized at the same time, it may be better to release them at one time since they are interdependent. Mirko said his concern is that four at once may be difficult to review.

**Action WGISS-38-18**: WGISS-All to provide comments to Data Preservation documents once they are sent out by Mirko.

### PV2013 Conference

Mirko Albani gave a summary of outcomes of PV2013, the 2013 conference for Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data (PV), held in November 2013. The objectives of the conference were to address prospects in the domain of scientific technical data preservation together with value adding to these data, providing a forum for organizations dealing with preservation of own data and value adding to present the status of their activities, plans and expectations, sharing knowledge, experiences and lessons learned, and fostering cooperation.

Mirko listed the conference topics:

* Data consolidation, archiving and preservation
* Metadata and standards
* Value adding and data exploitation
* Cooperation on data preservation
* Challenges to data preservation
* Interoperability and harmonization

Lessons learnt and future prospects and cooperation that resulted from the conference include:

1. Modular and flexible architectures for data preservation & access systems gives capability to change/adapt to different and evolving requirements; data archive management tools can facilitate the work of the data centres from various disciplines throughout a long period of time.
2. Diversity of instruments / data content and formats / metadata makes data preservation and knowledge representation very challenging (e.g. ISS); definition of “common core data models” and use of standards is key; special care should be considered for any data format to enable long term preservation and interpretation of data sets.
3. Common vision on data preservation among different organizations: data policy, common services, functionalities and processes. Definition and application of a common data policy (e.g. for historical data in Earth Observation, for ISS data) would be highly beneficial for data exploitation (and preservation).
4. Grid infrastructure and Cloud Computing can facilitate data preservation and exploitation but have limitations and can impose constraints; availability of shared infrastructure (e.g. at national level) and generic web services supporting basic functionalities common to most research data management systems can reduce overhead and costs for single organizations.
5. Big Data should not to be considered as a challenge but as an opportunity; Data Volume, Value, Variety, Velocity (4Vs).
6. Fundamental to share knowledge, experiences and lessons learned; cooperation, communication, awareness raising, interaction with users.
7. From LTDP to LTDP4V

Jonathon complimented Mirko on an excellent presentation.

### NASA Data Stewardship Activities

Andy gave presentation on behalf of Dr. H. K. Ramapriyan on NASA’s data stewardship activities. ESDIS is responsible for NASA Earth Science datasets many of which are over 20 years old. During the life of the mission, ESDIS actively manages all data in the collection, and has developed preservation techniques to ensure stewardship of the data at the end of the mission. NASA is continuing studies of stewardship and provenance at two external meetings.

NASA is not a permanent archive agency, but must maintain “research archive” for as long as data are used for scientific research and/or transition responsibility to permanent archives; NASA has to ensure data and other critical items are preserved and made available to permanent archival agencies. NASA has developed Preservation Content Specifications for Earth Science Data, and is participating in Earth Science Information Partners (ESIP) Data Stewardship Committee, on an “emerging” Provenance and Context Content Standard.

Andy listed the categories of content to be preserved, and listed the entities holding relevant content sometime during object/data lifecycle. He stated that ESDIS regularly reviews the status of the data collection to ensure safe-guarding of all datasets, and performs backup impact study showing degree of user impact resulting from data loss. He also listed the NASA working groups involved in data preservation, and described their activities with ESIP.

Richard asked if they also preserve software; Andy replied that they do, though it is not easy as operating systems change; they test the software every few years. Jonathon asked if they have a target of how quickly they can retrieve data. Andy said there is no exact target and Jonathon commented on the difference between theoretical reproducibility and practical reproducibility.

Kristi Kline asked if they have considered how the concept of Authoritative Data Source might be used within the various portals. For instance, if you search Landsat, you will get numerous results. How does a user know which source might be best to use?  Andy replied that for their own datasets they can trace the source, but for community archives it is another issue. Gabor also asked about traceability, and Andy replied that when they reprocess the data the popularity determines it. Mirko asked if their documents can be pushed to the DSIG, especially the Context and Provenance documents. Jonathon commented that the issue of preservation is becoming not just a technical issue but also a legal issue as the data is requested by the court systems.

### USGS Data Stewardship Activities

Kristi Kline reported on the USGS Media Study, which completes every two years, to investigate options for off-line archive storage. EROS maintains several versions of holdings: off-line and off-site in archive facility, near-line in robotic tape silo, and on-line disk (not all data sets).

Tape technology is used at EROS. 2014 technologies considered are Oracle T10000D, LTO6, and IBM TS1140. Kristi listed capacities of these, and Nitant asked if the capacity is compressed or un-compressed. Although IBM TS1140 scored highest, EROS intends to continue using LTO6 technology. TS1140 is not compatible with SL8500 near-line robotic tape library, and there are substantial costs to reconfigure for different technology.

The EROS film archive includes around 60,000 rolls of film. Of these 19,000 rolls are acetate, and this film is subject to vinegar syndrome, where the acetate film reacts to moisture causing the plastic base to bubble, warp and eventually become unusable. This decaying process emits a vapor similar to the smell of vinegar. Beginning in 1999 they began testing them with PH strips to determine their level of degradation. The most affected rolls were treated with molecular sieves that lessened the rate of deterioration. The worst rolls were copied to safety base film before the Photo Lab was closed in Jan 2005. Options to resolve V/S include modifications to the Phoenix systems. A science review culminated in the endorsement of 25 micron scanning to capture all of the scientific information from the V/S film.

Due to the EROS Photo Lab the shutdown, the LTA built the Phoenix IV systems to share imagery with science and develop a scanning strategy (2004 – 2010), and over 6.5 million aerial frames were digitized and released to EarthExplorer. The Phoenix systems were modified to accept 25 micron scanning backs (Phoenix V) that took the next step to scan and correct image products for long term archiving and release the film media to NARA (2010 – 2017) for preservation.

USGS now has 3-meter SRTM data, and been loaded to IDN.

Mirko said at ESA has the same problem with tape media converting to new technology. The difficulty is also the availability of the drivers. It is proceeding well, but requires a great deal of effort and time. Kristi added that the media is degraded to the state where they actually have to bake them at a low temperature. Gabor said GSDI has a MOU with ICA which has a working group on how to conserve this data, and the exchange of information would be beneficial to the community.

### JAXA Persistent Identifiers Study and Discussion

Yoshiyuki Kudo presented the JAXA Persistent Identifier (PI) study. He stated that the purpose is to obtain transparency of data use, which is increasing in importance as data becomes more easily accessible. The study has just begun, with identification of key elements that need to be addressed, and will take advantage of WGISS/DSIG best practices. The key elements are which persistent identifier to adopt, to which digital object to assign the PI, what should the syntax be, what the landing page look should like, which RA to choose, what changes are to be made in existing systems. Other elements are to determine the long-term commitment of maintaining the PIs, what changes to the everyday operation with the PIs, and what are the means for tracking use and statistical analysis.

To determine the digital object, JAXA will begin with clarifying what a data collection is.

For the landing page, there is a lack of consistent dataset collection documentation pages. Can the IDN dataset pages be used as a landing page? Agencies are already registering their dataset collections in the IDN, and it is robust, well-organized, and with consistent information. However, the monolinguality is a problem for global users. Michael Schick, Mirko and Andy commented on this, agreeing with the key elements. The IDN is looking at generating an automatically generated landing page depending on the metadata. Michael noted that for some organizations, the Digital Object Identifier (DOI) resolves the PI, so they can reference the dataset from there. The data centers are asking for a personalized landing page, but there is a need for consistency.

Alexey commented on handling translations within the GCMD. He suggested as a solution to narrow the problem down to looking at subsets of the keywords. If the agency could translate these as a first step, they can be can exposed, and the agency can be responsible for the translation. Nitant suggested associating the PIs to get PIs at a global level. Andy said there is a field in the DOI that agencies are strongly encouraged to use, as a prototype. Nitant suggested creating an automatic PI, but Andy said this may be an option that is too expensive. NASA is doing it for a new system (with 200 newly minted DOIs) and he will report the results in future. Richard said that CNES is using the same approach as EUMETSAT, managing them not just for EO data, but for all data. At the moment they use ARK instead of DOI. Michael said it is possible to put the ARK in the DOI - they are working to make this the entry ID. Richard said difficult when apparently same data with different processing chain, collection name is the same but PI is different.

Richard asked if there is a preference between DOI and PI. Jerome commented that with a randomly generated PID there is no way to know that it is a new version. But the decision is that the versioning should be in the landing page.

**Action WGISS-38-10**: Andy to review the use of the term DOI in the GCMD DIF metadata field to determine if it is appropriate to use PI instead.

### Persistent Identifiers Experience within LTDP WG

Katrin Molch gave a presentation on the PI experience within the Long Term Data Preservation (LTDP) Working Group.

The objectives of the PI Project are to assess if and how persistent identifiers can be introduced in Earth Sciences and to Earth Observation mission data; globally unique, unambiguous, and permanent identification of a digital object for locating and accessing it for a long time; evaluate harmonization / interoperability options.

Katrin reported that the PI preliminary study is complete, and pilot implementations (proof of concept) have begun. She listed as PID system requirements that they be globally unique, persistent, resolvable, reliable, authoritative, flexible, and interoperable. There are obvious benefits to users and data holders, and they are a shared responsibility between the PID system provider and the data holder.

The main outcome of the study is that Digital Object Identifiers (DOIs) are considered to be the most suitable approach for the use of Persistent Identifiers in the Earth Observation domain in Europe. It is preferred to associate PIDs to collections at a specific product processing level/product type and not to individual products/scenes, and to assign PIDs to Archival Information Packages (AIPs) and not to Dissemination Information Packages (DIPs). PIDs should be searchable in catalogues to reach data. PIDs should lead to a landing page where information on accessing the dataset is provided. Thus, there is no need to assign a new PID when the data location changes. PIDs should not be hierarchical. The relation between different datasets should be in the landing page. Provenance information should also be present in the landing page. In case of inter-agency data records, a PID (e.g. based on DOI) can be associated with the listing of all contributions and their respective PIDs (e.g. DOIs, ARK). They should provide recommendations on citation of data records.

PID Best Practices will be developed as a result of agreement on general aspects of PID implementation to achieve a degree of harmonization among European EO data holders.

Sample use case – data archive perspective: A data set series - which has a PID assigned to it – is reprocessed using an improved algorithm. This happens every few years. The data center is not planning on keeping more than the current and two previous versions of the data set series. The older versions will be deleted from the archive. Is the data center committed to keeping all versions? Does each new processed version get a new PID? How to deal with PIDs for deleted versions? The recommendation is that each reprocessed data set series receives a new PID. Ensure that the older versions are kept as long as possible, but definitely use a “tombstone” landing page which refers to the updated versions for any old versions that have been deleted.

Katrin also described the DLR Earth Observation Center PID activities and pilot implementation, and the SEASAT pilot implementation at ESA.

### Purge Alert Procedure and List

Mirko Albani discussed data loss prevention. He noted that EO data are episodic data and cannot be recorded again at some point in the future. They constitute a humankind asset which is fundamental in many Earth Science application domains and for scientific research. It is the responsibility of all agencies holding data to assess the relative value of their holdings and their requirements for long-term maintenance of those holdings. Sometimes an agency must make the decision to “purge” (i.e. to delete or remove EO data from an archive) data that could be important to the mission requirements of another agency. In these cases, defined and controlled procedures have to be in place to avoid, or minimize, the loss of EO data and to allow the handover of preservation responsibility to another agency.

The objective of the data purge alert procedure to prevent, or at least minimize, EO space data loss. It should be applied by all agencies holding Earth Observation space data in case, for whatever reason, they decide to no longer preserve one or more of their dataset. The process is to notify other archives of Earth Observation data holdings scheduled to be purged, and to offer these data to other archive centers. Mirko made the following proposal for the Purge Alert procedure, for WGISS to discuss and approve so we can move forward.

**STEP 1 – Purge Alert**

Organizations intending to dispose an EO Data Record should inform other organizations, before any purging of the data, through the “CEOS Purge Alert Service” with the goal to trigger a possible transfer of responsibility for the Data Record preservation to another interested entity. Purge Alert Service members can be contacted sending an email to purgealert@wgiss.ceos.org. When contacting the CEOS Purge Alert Service, the following information should be provided:

1. List and description of the Data Records to be purged and of the available associated knowledge (i.e. SW tools and documentation).
2. Data Records time span, geographical coverage, size, media of storage and archiving format.
3. Results and documentation of the Data Records Appraisal (e.g. <http://eros.usgs.gov/government/ratool/>)
4. Proposals to facilitate data transfer and ingestion to third party.

**STEP 2 - Purge alert response**

Organizations members of the Purge Alert Service whose missions match the type of Data Records being disposed, and who are interested to take on the responsibility for their preservation, should respond to the “Data Purge Alert” within two months from the reception to start negotiations with the alert initiator.

**STEP 3 – Archived data handover**

A Purge Alert initiator should provide support to interested organizations that responded to the “Data Purge Alert” to allow them to assess the value and prospects of the data set and to carry out the handover of the responsibility of their preservation.

Mirko displayed the purge alert page from the WGISS website, noting that it needs to be updated. The next steps are:

1. Draft, Review in DSIG and Issue the Purge Alert Procedure
2. Update Purge Alert Service web page content (<http://wgiss.ceos.org/purgealert/>) and migrate into the CEOS WGISS Web Site
3. Update the purge alert mailing list (purgealert@wgiss.ceos.org), selecting one contact point for each CEOS agency, or linking it to the WGISS-All mailing list. What is the WGISS recommended approach.

Andy asked if this is something other agencies commonly do and recommended a clear definition of what it means to purge. Mirko read some purge examples from the past to better understand what it means. Purge alert is only from the responsible (original copy) agency; if an agency is not the authoritative source they have the option of offering it up to anyone.

Richard suggested a fourth step: Obtaining from Plenary Agency Principals the correct contact person.

**Action WGISS-38-19**: DSIG to finalize the purge alert process using CEOS principles.

## CEOS Interoperability Interest Group

**Action WGISS-38-6**: WGISS Exec to identify a lead of the Interoperability Interest Group.

**Action WGISS-38-7**: Lead of Interoperability Interest Group to develop a page on the WGISS website.

### International Directory Network

Michael Morahan gave a report on the activities of the IDN Interest Group. Michael began with status of WGISS-37 actions:

WGISS-37-4: Michael Morahan to update the SEO’s spreadsheet of instrument/mission combinations noting which are available through the IDN on the Data Policy Portal. Action completed: IDN created a spreadsheet of all the IDN holdings (Data Sets) of CEOS members Platform/Instruments and delivered it to the SEO.

WGISS-37-11b: Michael Morahan to do the analysis of the ISO to DIF converter on the five examples from ISRO. Action completed: ISRO delivered DIFs to the IDN and they were reviewed and loaded into the IDN. Michael listed the ISRO records.

Michael gave IDN Metrics for the past 12 months; total visits to the IDN site is close to 30,000, and total pages viewed close to 60,000. He listed the growth of IDN metadata records. A chart of CEOS DIFs shows population by month, with a peak from the entry of the Antarctic DIFs. Michael showed the number of IDN records updated by month, and the DIF breakdown of the GCMD (which is the overall system – IDN is just NASA contribution to CEOS). The CEOS entries are the bulk of the entries to the GCMD, and second is NASA, and only 15% from others. Andy commented that it would be interesting to track the trend over time. IDN usage by continent and country was displayed, showing that Asia is the leading viewer of the IDN. Australia is last but is the largest contributor with the Antarctic entries. Nitant suggested changing the label to IDN access, and to differentiate between countries contributing and countries accessing the IDN.

Michael displayed user access of CSW and GCMD API services, CWIC datasets by topic, US GEO/GEOSS metrics, current IDN/GCMD database statistics, and US GEO Data Core contributions by agency. An IDN statistics page is available on the IDN web page, as is contact information.

Michael announced that they are going to start sending out more proposals for additions to the IDN, and they welcome suggestions and feedback.

Yonsook commented that the CWIC team will be contacting the IDN team to work with Alexey to create DIFs for the Russian data. Yoshiyuki asked if the ranking he mentioned affects the order of returned datasets from searches. Andy said he had a discussion with Chris Lynnes about this. It was agreed that it would be advisable to have a discussion in WGISS about ways to enhance results page (returned datasets). Yoshiyuki said that would be a very good input for the OpenSearch project.

Andy Mitchell explained new features of the IDN:

 Common Metadata Repository (CMR) GCMD/IDN support for the CMR.

 Metadata review and assessment, tool development, outreach documentation, UMM-C metadata model.

NASA rolled out the first part of system in August, and one of the big improvements is performance.

The new INTEROP proposal for new platform/instrument/sensor proposes The "Platform (Source\_Name)" and “Instrument (Sensor\_Name)” fields identify which satellite, instrument and/or sensor is used in the production of the dataset. The proposed action is to move the fields into one structure group. This format can help the users refine their search quickly using the related platform/instrument/ sensor keywords associated with a data set. The order of the field will be a drill-down format starting with the platform name then associated instrument(s) and finally the associated sensor(s) of the instrument. The sensor, platform name and the instrument will be required; the rest are optional. The group will be repeatable and also repeatable within the sub-fields. Keywords must be selected from the Controlled Keyword List.

The metadata tagging table is now operational. It allows partners to tag metadata with arbitrary "flags” to organize collections of metadata to a specialized discipline or interest outside of what is defined in the metadata. Tags are not publicly displayed in any web interface, and can be used in creating portal subsets or any other searches. Now promoted to Production (gcmdservices) and currently being populated by CWIC. The metadata mapping table will have an automated feature to populate the CSW server entries. The script will automatically populate the CWIC tag CWIC tag in the TEST and Production CSW servers.

The IDN progressing from user interface system to server interface system.

The Metadata quality assurance service, or just QA Service is a new proposed system. There will be a science coordinator, to do a thorough review with data provider for keyword checking. There will also be automatic systems that can fail, and this will send a notification. Andy said when metadata gets stale they hope to automatically detect that. The CMR process has rubrics where it scores the data in the docBuilder, and checks that values are there and also that they are sensible. Examples were shown.

### Federated Earth Observation (FedEO)

Mirko Albani gave a presentation describing the Federated Earth Observation (FedEO). He explained that FedEO provides brokered discovery, access and ordering capability to European and Canadian EO missions. FedEO implements the OpenSearch OGC (and other) interfaces for an increased number of discoverable and accessible EO data collections, and for interfacing with CEOS Community Catalogues and Clients. He showed a diagram of the FedEO brokered architecture diagram displayed, and indicated that any client to OpenSearch can access FedEO. He listed the backend connectors supported by FedEO.

#### Status

Mirko reported the current status: FedEO end-point/component is being migrated to ESRIN, and completion is expected by end October 2014. M2M External access tests with GEO Discovery and Access Broker (DAB) completed and access implemented. FedEO is accessible for demonstrations and administration through a dedicated Client Portal: <http://services-test.eoportal.org/web/guest/fedeo-demo-services>. Information on OpenSearch implementation available at: http://fedeo.esa.int/opensearch/readme.html (Main), ttp://geo.spacebel.be/opensearch/readme.html (Test). Further catalogues interfacing ongoing (e.g. DLR, EUMETSAT, VITO, NASA ECHO, Disaster Charter, SPIRIT).

As of September 2014, 511 collections are discoverable and accessible via FedEO, plus the CWIC collections.

Mirko showed a video presentation demonstrating the FedEO catalog two-step search. Andy asked if CWIC is going to be a catalog of FedEO. Mirko replied that the user can select a catalog from which to filter, or can search from all. Yves said the user can control the search based on which parameter is chosen. Andy asked if that is the same functionality that CWIC will have. Yonsook said that yes, if there is a FedEO connector. If everyone has OpenSearch implemented then in principle it can be done. If there are OpenSearch systems outside of CWIC and FedEO, the client needs to know it.

Yves said the metadata is all in ISO, and Yonsook said Michael Morahan can do a bulk upload to manage the process to load to the IDN. Yves said they sent a file sample to test. Michael Schick added that EUMETSAT did it and it was very simple for them (they are already Fedeo compliant).

#### Partner Guides

Yves Coene continued the presentation, discussing the FedEO partner guides. The Client Partner Guide contains a general description of the FedEO environment and its relation to other systems. He also described the Data Partners Short Guide, explaining how to be accessed by FedEO.

Step-1 - Implement one of the standard interfaces (for dataset search).

Step-2 – Send endpoint to ESA FedEO team for integration.

Step-3 – Provide ISO 19139 Metadata file for each of the proposed collection identifiers EOP:MYORGANISATION:XXX-YYY including temporal and geographical coverage. Can be created with support of the ESA FEDEO team (e.g. through answers to a questionnaire). Sample metadata files can be provided as well.

Step-4 – ESA FEDEO team provides test endpoint (OpenSearch) for organisation XXX to preview their OpenSearch I/F. Optional steps: agree keywords to be added to collection metadata as per agreed thesaurus, and federated authentication (SSO).

Currently OpenSearch access to HMA Catalogs is supporting OGC 06-131, CWIC catalogs, Virtual Archive 4 and G-POD. OGC 10-032r8 Geo and time Extension, and OGC 13-026r4 Extension for Earth Observation are implemented.

#### Demonstration

Yves demonstrated FedEO showing collection search and product search. He demonstrated the two step search approach (search dataset series and then search datasets inside the selected dataset series).

Yonsook asked how many of the collections are GEOSS DataCORE. Mirko said they are still clarifying this with the providers. Michael Schick said they have extended their collections to mark them as DataCORE but have to go through all the collections.

#### Future Work

Future work is to populate dataset series catalogue with additional dataset series metadata. Currently, not all dataset series have ISO19139 metadata, so two-step search not yet supported. They need to integrate additional back-ends and perform interface improvements: Add RESTful interface in addition to SRU-style interface, support content negotiation (via HTTP header), and support additional queryables at collection level, and align with CEOS Best Practice and INSPIRE Technical Guidance (Download Service), integration with CNES RESTo Query Analyzer and Web browser search box. In addition, open Linked Data, additional media types (HTML media type with schema.org annotations and RDF media type for dataset series), and make available FedEO metadata to the Linked Open Data (LOD) community.

### CEOS WGISS Integrated Catalog (CWIC)

#### CWIC Report

Yonsook Enloe gave a report on the CEOS WGISS Integrated Catalog project. She began describing CWIC basics, architecture, data partners and data, data partner responsibilities, client partners, and support for new partners. She listed the data partners, adding the following new partners:

* ISRO will be operational in the next month.
* EUMETSAT will become operational in the next several months.
* ROSCOSMOS to start work on their connections.

RADARSAT-2 data is also now searchable.

Yonsook reported on progress of CWIC toward OpenSearch. A broad review of the OGC Geospatial and Temporal Extensions and OGC Extensions for EO documents in summer 2013 with ESIP OpenSearch implementers; comments were given to the CEOS OpenSearch team who submitted it to ESA. Most comments incorporated into the revised OGC specification were released in late fall/winter 2013. Ongoing activities are bi-weekly teleconferences with the CWIC, IDN, and CWIC-Smart teams, discussing an overall design for the OpenSearch implementation, identifying design/implementation issues, and coming to resolution on these issues. In January 2014 there was a CWIC team meeting, focused on making decisions on OpenSearch implementation issues and firming up the CWIC OpenSearch Best Practices.

Yonsook reported for end March/early April – Early prototype implementations was made accessible for team testing. Summer 2014 – detailed comparison of CWIC, FedEO, and CNES OpenSearch implementations, and came to agreement on the minor differences. A September 2014 teleconference confirmed the agreements. September 2014 – CEOS OpenSearch Best Practices, draft document released to WGISS for review. There may need to be final revisions to implementations – to conform to CEOS OpenSearch Best Practices. The CWIC-Smart Client that supports OpenSearch is available.

CWIC Servers (DEV, TEST, PRODUCTION, STATIC) working well on a new hardware platform at USGS/Sioux Falls since summer 2013. The CWIC servers at GMU have been shut down. The team is solidifying test plans and processes to support new CWIC server enhancements, new data partner addition, IDN/CWIC synchronization, data accessibility testing (new data, confirmation of existing data); automating testing wherever possible. Enhanced metrics collection and display are being implemented, and the IDN/CWIC synchronization plan is simplified with the development of the IDN “tagging”/mapping tables. The metadata about metadata is kept separately as “tags” or mapping table. Example: Project=CWIC is now a CWIC tag. This will enable clients to tag a subset of CWIC data for search and access. They have updated the CWIC Data Partner Guide, the Client Partner Guide, and Exceptions Handling document to reflect OpenSearch support. ISRO will become operational in the next couple of months, EUMETSAT will become an operational data partner in the next several months after decisions made on what data to make accessible via CWIC, and they are starting discussions on the Roscosmos CWIC Connector.

CWIC Future work includes work with the CEOS OpenSearch team to come to agreement on the CEOS OpenSearch Best Practices and adjust implementations to conform to the document. Interoperability testing with the FedEO OpenSearch implementation and study of the feasibility of FedEO Connector to CWIC are also planned. The CWIC Data Partner Guide, CWIC Client Partner Guide for CEOS OpenSearch will be updated. Support the developing CWIC Data Partners to become operational is an ongoing activity, as is strategy planning for supporting VCs, WGD, SDMS, etc. for data access issues. CWIC will continue to support GCI integration, and develop a strategy for “Service Level Agreements” – how to get agreements on improving the reliability of the CWIC service, including agreements with data partners.

Both the IDN and CWIC are integrated with GEO, and accessed by the GEO components (Geo Web Portal, DAB, and GENESI). IDN provides collection metadata for over 24,700+ data collections available to GEOSS (April 2014), with 9600+ of these data collections tagged as “geossDataCore”. Registering a data collection in the IDN in effect registers the data collection in GEOSS. A data collection must be registered in the IDN to be CWIC accessible. CWIC provides access to over 1900+ (Sept 2014) data collections with inventories (over 70 million granules, adding granules daily from live missions), with all but 2 collections “geossDataCore”. GEO Director, Barb Ryan, noted during the GEO session at the 2013 CEOS Plenary that of the 65 million granules accessible in GEO (Oct 2013), CWIC provided access to over 50 million granules. The team will send CWIC metrics to GEO Secretariat on regular basis.

Information and resources can be found on the CWIC page on the WGISS site.

#### CWIC Server Implementation

Lingjun Kang gave a presentation on the CWIC server implementation. He described the CWIC CSW and CWIC OpenSearch protocols and displayed a diagram of the CWIC CSW protocol adaptation and the CWIC OpenSearch protocol adaptation, and explained how identifying the data provider from IDN Entry ID works. The CWIC client searches for granules from CWIC using the IDN dataset Entry ID to specify the data set. CWIC identifies the data provider to which the request is to be dispatched based on IDN dataset Entry ID. CWIC uses the data provider’s native dataset ID in the granule search request, not necessarily the IDN dataset Entry ID. Catalog ID is identified for data provider dispatching, and native dataset ID is also identified for constructing a query with the native dataset ID recognized by data provider inventory system. The information in the internal mapping list is kept updated through daily synchronization from IDN.

Data provider protocol and requirements are as follows: As a mediator, CWIC’s capabilities of querying and returning response rely exclusively on the native protocol supported by data providers. To be better integrated with CWIC, data provider inventory systems are expected to meet the following requirements: Query by dataset identifier, spatial/temporal query, support pagination, support returning essential metadata info, and exception handling. Requirements of query interface are: Support query by dataset identifier (native or Entry ID), support spatial query by bounding box, support temporal query by start/end date time (or either). Requirements of pagination support: Support query by cursor (i.e. start position, maximum returned records, next record); indicate pagination info in query response (i.e. number of returned records), essential metadata info are expected in native metadata of data providers was listed. Requirements of exception handling: Data provider remote inventory system should indicate errors and return exception status through either application level (e.g. CSW exception response) or generic protocol level (e.g. HTTP status code).

Archie Warnock described the server architecture for Multiple Request/Response Formats with a descriptive diagram.

Kang continued with unified error handling: CWIC exception consists of exception ID, message and HTTP status code (only available for CW IC OpenSearch). Exception examples were listed. The process for exception handling is: Collecting exceptions locally from CWIC or parsing exception message from remote inventory systems; Map to proper CWIC exception; Connectors pass exceptions back to Mediator; Mediator communicates exceptions to clients in standard ways at both transportation (HTTP) and application (message) levels; Error text reported back to the client is added by the Mediator, not by the Connectors, in order to provide consistency. Communicate error message at application (message) level: Return exception message through CSW; Return exception message through OpenSearch; Communicate error message at transportation level; Indicate exception code in HTTP (only CWIC OpenSearch).

CWIC Metrics: CWIC metrics collect essential information from application log, including: Client information (client IP and client id in OpenSearch metrics); Transaction information: query string, request type (CSW), requested output schema (CSW), requested element set name (CSW) and query start/end time; Queried resource information: dataset ID and granule ID. CWIC metrics are analyzed periodically to provide the following reports: Client summary (by IP, client id, client type and country); Daily query auditing; Visit auditing by dataset (e.g. most popular dataset); Visit auditing by data provider.

Alexey asked about bounding box size. It is dependent on the data provider; if there is a limitation, the information can be sent back to the user. In response to a question about error handling, Kang said that from the reference documents they will be able to see the error and present it in a way that makes sense to them. This is an area where conformance is important. Andy commented on the metrics – can they indicate who is adopting OpenSearch.

#### CWIC/IDN Synchronization through Metadata Mapping Table

Michael Morahan explained the metadata mapping table feature which allows partners to tag metadata records with arbitrary "flags” to organize collections of metadata to a specialized discipline or interest outside of what is defined in the metadata. Tags are not publicly displayed in any web interface. Tags can be used in creating portal subsets or any other searches. The feature is currently available and is already being populated by CWIC and IDN.

He listed the recommended fields. He explained that a general GET output with no Entry\_id will return a list of all tags and documents. One document is returned, if an entry\_id is supplied. By default, entry\_id and Extended\_Metadata are returned. Optionally additional metadata information can be specified in the "sections" list and these tags will be included in the output XML (example: Spatial coverage and temporal coverage).

Andy asked about tags that are not publicly displayed – Michael replied that it is up to the client. Entries can be removed from the table (not the DIF) only by those who have access; this can be used to temporarily disable an entry. Yves asked if you search IDN can you get ISO metadata. Michael replied that yes you can. Yoshi said the water portal plans to use the mapping table. Initial plan is to use IDN for the collection level search; the Water Portal will not use the IDN because of complication with coordination with providers of water data.

Nitant said mapping table helps identifying the assets of different organizations. What are the tags that are used? Users can create their own groups.

#### ISRO Partner Status/Demonstration

Nitant Dube gave a brief overview of the ISRO Earth Observation Catalogue System (IEOCS), and described the metadata available from ISRO connectors, listing MOSDAC and NRSC connectors.

Nitant reported that the ISRO-CSW setup is complete, the ISRO-CSW-MOSDAC connector is in place, the INSAT-3D DIFS (14) are public, and the ISRO-CSW integrated with CWIC-CSW and CWIC-OpenSearch Connectors (in TEST setup). He listed the 14 INSAT-3D Imager public DIFs, and provided the demonstration link <http://api-test.echo.nasa.gov/cwic-start>.

Future plans:

* Operational release of direct ordering from MOSDAC
* Registration of remaining DIFS
* Integration of NRSC connector

Nitant gave a quick demonstration, and have implemented direct order capability.

Yonsook said that is very exciting to see.

#### EUMETSAT Partner Status/Demonstration

Michael Schick gave a presentation on the EUMETSAT implementation.

Michael explained CWIC in the context of the EUMETSAT implementation. EUMETSAT’s ‘EOPOS Adaptor’ is an implementation of OSCWIC and EOPOS. CWIC uses a wrapper with a connector implementing a client of the OSCWIC interface (OSCWIC Connector). EUMETSAT EOPOS Adaptor implements the OSCWIC Interface. EO Portal will become ESIP (CWIC Earth Science Information Partner) via this Adaptor. He described the computational view of OSCWIC. The workflow for the EOPOS Adaptor is: SCWIC Connector will request the OSCWIC OpenSearch Document (OSDD). Search requests can then be formed as defined by the OSCWIC Specification and the OSDD (including the collection identifier from GCMD). Responses as defined by the OSCWIC and EOPOS Specifications. EOPOS Adaptor maps request to U-MARF Interface. This is executed on U-MARF catalogue. UMARF response is converted to EOPOS syntax. Response returned to originating OSCWIC client.

Michael also described the CWIC engineering/technology view.

In terms of status, a few sample collections are configured in the moment. The will continue to validate/test different type of satellites and data: GEO data (MFG, MSG) and LEO data (MetOp, NOAA).

Access to data is still to be completed. Capability exists for guest download or registered EO Portal user. Guest works with E-mail address check. Agreement of Collection is available via CWIC though internal discussion and approval required. Implementation of Dataset Access with click request and download is possible by March 2015. Presently the system is deployed on Validation and Test platforms. Transition to Operations requires Joined Operations ICD (JOICD) with contact points.

Yonsook commented that they have really enjoyed working with EUMETSAT; it was a fast implementation, especially since EUMETSAT had previous experience with HMA.

#### AOE Partner Status/Demonstration

Chaoliang Wang reported that the AOE CWIC team mainly maintains the Chinese node system and is keeping it in operation. The have also tested the connection between the Chinese node system and the CWIC server with the help of Kang and it is working well.

Yonsook said thank you, the team is ready to assist in way they can.

#### Roscosmos Partner Status/Demonstration

Ovnan Tokhiyan gave a presentation on the IGDIS ERS current state and prospects of development. The project purpose is integration ERS information resources into a unified GEO-information space: providing ERS data for end users. Tasks include orbital constellation of ERS satellites application planning, IGDIS ERS ground-based infrastructure application planning, data reception and processing from Russian and foreign ERS satellites, classification and storage of ERS products, ERS products unified cataloging, and access to IGDIS ERS information resources via geoportals and web-services.

Ovnan displayed a map showing reception and ERS data processing centers, and a diagram of the data flow and of informational systems that use web services of Roscosmos Geoportal. The direction of development is enhancement of data reception from ERS satellites, processing capabilities enhancement, and enhancement of existing infrastructure for ERS data distribution – providing wide range of ERS informational products, and creating cloud computing data processing centers in IGDIS ERS centers.

Limiting factors are the variety of technologies for ERS product creation are insufficient provision of input ERS data, insufficient provision of computational resources, and lack of unified access tools. Placing of ERS products technologies in cloud infrastructure of IGDIS ERRS data processing centers. Interaction with integrated information system for basic space research of the Russian Academy of Sciences. Integration with base products bank, and with CWIC.

Alexey Gladkov continued with a presentation of the Roscosmos Geoportal, which is the main geo-information service of Integrated Geographically Distributed Information System of ERS (IGDIS ERS) for web application and web services. Resources available are integrated catalogue of IGDIS ERS, and GeoData bank of IGDIS ERS. Alexey listed the information layers for imagery, and described the connection with CWIC, with ROSCOSMOS GeoPortal as the data provider.

Main features of the connection with CWIC are authorization, HTTP access, and JSON or XML responses. Roscosmos Geoportal provides the native search service. Available functions are authentication, requesting data from classifiers (platforms, instruments, sensors, processing levels, etc.), satellite imageries search, and receiving metadata in format of ISO 19115 profile.

Alexey summarized the activities: 1. Register Russian Platforms/Instruments/Sensors in the IDN 2. Create metadata for collections (ISO 19115 format) 3. Update metadata for granules with required fields (type of metadata: for collection or granule and browse/order URL) 4. Translate some metadata fields those are required it 5. Create connector from CWIC to native Roscosmos Geoportal search service 6. Tests

He displayed a few examples.

Yonsook thanked Alexey and Ovnan and looks forward to working with them. It is great to hear that the metadata will be available, and orderable.

#### GEO Community Portals

Ken McDonald gave a presentation on the GEO community portals. He began with a reminder that the GEOSS Common Infrastructure (GCI) supports access to all registered GEOSS resources, and the GEOSS Portal interfaces with the DAB to discover, search and access resources in the GCI registries and those within larger federated EO catalogs. He explained the rationale of community portals, and continued with a review of the Community Portal Paper. The purpose of the paper is to provide guidance to developers of Community Portals, working with communities to identify and characterize functions and use cases of community portals, demonstrating the benefits of utilizing GCI, instructing on its use, and sharing experience of previous efforts (e.g. AIPs). The goal is to maximize benefits of GCI and promote community contributions to system of systems. He listed the team members.

The paper was originally limited to Community Portals/Clients, but the scope broadened to look at range of possible interactions between GCI and Community Infrastructures. This allows GCI to better leverage community capabilities, and promotes closer engagement of user communities in sustaining and enhancing GEOSS.

Ken gave an outline of the paper, and listed the release plan.

In summary, the “System of Systems” has tremendous potential to leverage capabilities of broad set of communities, to support general data discovery/search/access and to promote innovation/specialized services. It also has challenges: Components/interfaces evolve, leveraging introduces interdependencies, and there is a diverse sets of users. Communication between component, data and service providers and systems implementers/operators and users is key. Hopefully the document is one way to help.

### CEOS OpenSearch Project

Yoshiyuki Kudo introduced the CEOS OpenSearch Project session.

#### Status Update

Yoshiyuki gave the status of the WGISS OpenSearch project, whose goal is to establish an OpenSearch best practice of CEOS catalogs, including IDN, CWIC, FedEO, and other CEOS catalogs. The draft version of the CEOS OpenSearch Best Practice Document is complete. It contains 17 best practices, based on implementations in CEOS agencies, and classified as required, recommended, or optional.

CEOS comments for the OpenSearch were submitted to OGC. Those that did not fit in the OGC specification are the ones included in the Best Practice document. A few items in the draft are still to be resolved.

**Action WGISS-38-17:** WGISS-All to provide comments on the draft OpenSearch Best Practice document by end of November.

#### Towards Best Practice Version 1.0

Yoshiyuki stated that Version 1.0 completes the project unless additional activities are identified. To be completed yet is the resolution of comments and discussion items, and to rank the practices. It is expected that after release as a CEOS document, it will be visible to relevant people. How to best advertise it for proper visibility is up for discussion. Andy said a lot of this needs to be tackled inside the Interoperability Interest Group as it relates to other activities as well. It is also important to advertise this to other CEOS agencies not represented in WGISS.

Satoko reminded that there is an action for CEOS to develop a document repository. Andy said they are working on a new document management system, but this would not necessarily provided the desired advertising. Kim Keith is looking for inputs; perhaps a short spot on the CEOS page would be helpful.

Yonsook suggested having a specific time period for the comments towards version 1.0. The opportunity should be given with a limited time period for comments. She suggested finding a way to entice agencies to take a look at is as they all have catalogs.

**Action WGISS-38-4**: WISP to organize a teleconference to gather thoughts on best way to advertise important documents or recent document releases on CEOS website, and discuss with Kim Keith at the CEOS Plenary.

#### FedEO OpenSearch Implementation

Yves Coene gave presentation on the OpenSearch status of FedEO. He reported that the FedEO team has developed a client partner guide, containing a general description of the FedEO environment and its relation to other systems, with details about the FedEO query interface based on OpenSearch. It provides a description of two cases: one starting from a dataset series catalogue, and a second one accessing immediately the dataset catalogue. Details on the catalogue connectors and how their corresponding dataset series and dataset metadata can be discovered are also included.

Currently the FedEO OpenSearch access is to HMA Catalogs supporting OGC 06-131: Atom with EOP O&M (OGC 10-157) metadata as foreign markup or atom:link; CWIC catalogs: Atom with DC or ISO metadata as foreign markup or atom:link; and OpenSearch catalogs e.g. Virtual Archive 4 and G-POD. The interface aligned with OASIS searchRetrieve 1.0 conventions. Additional RESTful interface being prepared. OGC 10-032r8 Geo and time Extension and OGC 13-026r4 Extension for Earth Observation are implemented.

Yves described the FedEO OpenSearch architecture, including the two step search approach. The collection level search allows for same collection identifier in two catalogs. Yves gave an OpenSearch example, showing the request URL.

Work performed so far includes interface changes such as Return Simple GEORSS (CEOS Recommendation), Support for StartPage queryable, Support for eo:platform and eo:instrument in Collection level search (EOP:ESA:FEDEO:COLLECTIONS), and preserve semantic annotations in ISO collection metadata in Atom response (as per OGC 08-167r2). They have also improved Atom response in case of ISO collection metadata with semantic annotations (OGC 08-167r2).

Additional collections include NASA ECHO, CNES SPIRIT, Disasters Charter catalogs, and ESA DISSHARM – Landsat 5 etc.

Compliance tests are underway.

Yves listed opensource OpenSearch software, and gave the link for the components. Documentation available on ESA HMA wiki.

Planned future work includes query analyzer integration with CNES RESTo Query Analyzer, and simplified OSDD understood by Web Browser search interface.

Future work for the interface includes implementing CEOS OpenSearch recommendations, improving compliance with INSPIRE
Download Service Technical Guidance, and harmonizing Atom response messages from different back-ends. In terms of harmonization, they plan to harmonize Collection Metadata format (OGC 11-035r1) and Product Metadata format (OGC 10-157r4).

Open Linked Data will include additional media types: HTML media type with schema.org annotations, and RDF media type for dataset series, and make available FedEO metadata to the Linked Open Data (LOD) community.

Andy commented that he is glad to see the slide on the opensource. Yoshiyuki said he would add it to the opensource page on the WGISS site.

#### CNES OpenSearch Implementation

Jerome Gasperi gave a presentation on the OpenSearch implementation at CNES. The THEIA Land Data Centre is a national inter-agency organization designed to foster the use of images coming from the space observation of land surfaces. It offers the national and international scientiﬁc communities as well as the public policy for the monitoring and management of environmental resources, a broad range of images at different scales, methods and services related to the land surface observation from space.

Plateforme d’Exploitation des Produits Sentinels (PEPS) is the French Sentinels Collaborative Ground Segment. It will provide Sentinel 1, 2, 3 and 5P products. Data access to THEIA and PEPS is done through OpenSearch. It will align to CEOS OpenSearch Best Practices document by mid November 2015.

Yonsook said that the CEOS OpenSearch project has been a great project, allowing them to identify and discuss issues, and change implementations accordingly; it is a great example of collaborative effort. It has been a testbed effort – resulting in a long term establishment. Yoshiyuki has done a great job of putting together the documentation.

# CEOS OpenSearch Workshop

Yoshiyuki Kudo and Jerome Gasperi introduced the CEOS OpenSearch Workshop, which mainly targeted for implementers (engineers) of CEOS agencies. The objective is to learn to access and make better use of CEOS catalogs using OpenSearch, and to align OpenSearch server implementations with CEOS Best Practice.

## OpenSearch Characteristics

Yoshiyuki noted that OpenSearch was originally conceived by Amazon A9 for mass-market technologies for discovery on the Web. It has a few simple rules. Service description can be found in the OpenSearch Description Document (OSDD) for client application (URL template). Search can be done with a simple URL (HTTP/GET). Search response can be any existing format (Atom etc.) with a few OpenSearch parameters inserted. The search request is a basic html where specific parameter strings need to be replaced with the parameter values. The OpenSearch response: Query element describes the query used.

OpenSearch does not mandate a speciﬁc format; it enriches feeds in html, XML or JSON formats. Atom is widely compatible because it is foreign element friendly and is ideal to wrap and link resources. Atom entries should not be considered all-purpose envelopes (entries link to other resources representations). Atom is the middle ground that let mainstream application enter OGC world while linking more complex resources.

The OpenSearch workflow is as follows: Discovery (is OpenSearch supported?) Once the client finds that the server supports the OpenSearch, it identifies which URL template to use, builds the URL with search parameters, and returns the search result.

Search specifications are OpenSearch, OpenSearch Parameter Extension, OpenSearch GEO and time extension, OpenSearch EO extension, and O&M EO metadata model.

Nitant asked which of these are the minimum requirement for CEOS implementation. Yoshiyuki replied that it is the first three (OpenSearch, OpenSearch Parameter Extension, OpenSearch GEO and time extension).

## CEOS OpenSearch Best Practice

Yoshiyuki commented that speciﬁcation deﬁnes constraints and options. Best Practice provides hands-on implementation know-how. The purpose of the best practice is to promote the use of the OpenSearch standard as a means of data discovery for Earth Data providers, to deﬁne the expectations and requirements of candidate OpenSearch implementations, to remove ambiguity in implementation where possible, and to facilitate the aggregation of results between disparate Earth Data providers via OpenSearch common standards. It allows for clients to access search engines via an OpenSearch Description Document (OSDD) with no a priori knowledge of the interface, and facilitate smooth integration between related OpenSearch implementations, such as a dataset resource collection that refers to granule resource collections from another provider.

Prerequisites: CEOS OpenSearch Best Practice is compliant to OGC OpenSearch speciﬁcations, OpenSearch Geo and Time Extensions [10-032r8], OpenSearch EO Extension [13-026r4] (draft).

Yoshiyuki showed the outline of the Best Practice document, and listed the 17 best practices.

The first is the two-step search. This is a requirement due to satellite data granularity. He noted variation in terminology: a collection is also known as dataset, dataset series, or product, and consists of granules (also known as dataset, product, and scene). Andy suggested using ISO terminology.

Step 1 (collection level): Request for OSDD 🡪 search request 🡪 search result🡪 identify OSDD URL for a collection.

Step 2 (granule level): Request for OSDD on a specific collection 🡪 search result.

Support for OpenSearch parameter extension (recommended). The parameter extension explicitly advertises the valid lists and ranges of search parameters. This dramatically reduces ambiguity and errors.

Recommended search parameter support is required.

Best Practice 11: Metadata in search result. This is for detailed representation of the metadata, and for readability.

Andy said if schema.org is supported then the DIFs should be shown– at the last teleconference Doug Newman mentioned that Google has begun finding this.

Yoshiyuki only covered some of the Best Practices and recommended that the detail in document be examined.

## OpenSearch Implementations and Demonstrations

### IDN/CWIC

Doug Newman gave a presentation on the IDN/CWIC OpenSearch implementation. He reiterated that OpenSearch is lightweight and simple, standards-based, RESTful, low entry cost; FreeText+spatial+temporal satisfies most user needs, and the two-step querying provides higher precision. He displayed a diagram showing two-step searching diagram and gave an example of CwicSmart+IDN OpenSearch+CWIC OpenSearch.

Doug listed a few CWIC Best Practices, such as conforming to OpenSearch and OGC specifications, and adding the following recommendations such as make best attempt at coercing clients to provide client ids, extend ESIP ‘two step’ idea to link IDN dataset -> CWIC granule searching, embrace hypermedia – HATEOAS, relevance, utilizing the parameter extension for better clarity of service, and spatial (MBR) and temporal (Dublin Core) extent representation.

CwicSmart is a programmatic OpenSearch client. Any OpenSearch implementation that conforms to the CWIC Best Practices can be queried using CwicSmart.

For dataset search they use the OSDD to dynamically render an html form. Using the template and the parameter extensions they can determine whether a parameter is mandatory, its type, its bounds and whether it is a ‘core’ parameter like searchTerms.

1. The user submits the form and the results are presented
2. HATEOAS is used to generate links for traversing through the results. NEW: In the absence of such links elements mandated by the base OS specification are used.
3. Search links are used within the entries to provide ‘Search this <resource>’ links that perform two-step searches.

Granule search:

1. Parse the OSDD to generate an HTML form to enable the user to perform a search against a child resource (in this case granules pertaining to the dataset they are searching on).
2. All shared parameters are propagated to this search form. For example, bounding box. The OSDDs tell us whether they are ‘the same’
3. The user submits the form and the results are presented.
4. HATEOAS is used to generate links for traversing through the results.
5. Icon links are used within the entries to render browse thumbnails to the user
6. Enclosure links are used within the entries to allow the user to download a resource (granule in this case)

In the context of CWIC and OpenSearch the IDN serves as an OpenSearch implementation that provides access to collection metadata at an international level, linkage to CWIC granule discovery, and rich and canonical collection metadata. Doug explained searching, and search parameters, and described the CWIC best practices, and parameter extension.

In the context of CWIC and OpenSearch, CWIC serves as an OpenSearch implementation that provides access to granule-level metadata and data from disparate international providers, and uniform access to that metadata and data. He displayed architecture, searching, search parameters, and best practices, and gave examples of OSDD response, and the granule search request and response

In conclusion this is a widely adopted standard that is extremely easy to use. Implementation time of OpenSearch API or client is short (two months). Linkage between data providers is easy to achieve, and CWIC OpenSearch best practices are robust as proven by CWICSmart

What next? Aligning with CEOS Best Practices, getting other APIs working with CwicSmart, increasing granule precision by temporal and spatial ranking, facetted search, and substituting discovery with processing. It is also desirable to control the problem of when someone selects a large temporal or spatial range.

### GCMD/CWIC

Thomas Cherry gave a presentation on the GCMD-CWIC OpenSearch. He began with an overview of the functionality, and said they have added client ID metrics, optional secondary OSD, parameter extension, and improved pagination support.

The client ID is incorporated in the URL and passed along. He showed an example of the OpenSearch parameter extension. CWIC compliant requests standardizing around a common set of pagination parameter names with results all producing navigation link tags. Thomas gave a demonstration.

Michael asked him to talk about how long it took him to create the GUI. Thomas replied that it was no more than two days, and that includes learning a new technology at the same time.

### FedEO

Yves Coene gave a demonstration of FedEO, noting that the OpenSearch Description Document (OSDD) is the start of everything. He copied in the URL which returns multiple bounding boxes and a result set with browse image. Since not everyone supports radial search the user can go to the center and make a rectangular bounding box.

It would be good to have Best Practices on what should happen with unexpected conditions. They have documented handling these in the exception guidelines handbook. This ensures consistency across the data centers. It would make sense within WGISS to discuss these conditions and default behavior. These are key things that make the system work well. Have to have agreements when there are multiple data centers.

### CNES

Jerome Gasperi gave a demonstration of RESTo v2 – Open source OpenSearch server for EO data: Search, visualize and download EO data: Restful, Responsive, and Reliable. He described the architecture. The Abstract Database Access Layer is PostgreSQL Driver. Inside the database are two schemas, the RESTo, and user management. The Search for data is in the database, as is the ingestion. During the ingestion process, resources are automatically tagged thanks to iTag library (standalone library). iTag provides semantic enhancement of EO data.

Michael asked how to deal with if you decide to extend. Jerome from the library you can easily add after original ingestion.

Out of the box tagging sources are continents, countries, regions, states, cities, land cover, population count. Jerome displayed an example, noting that it is superfast because there is no keyword search. Everything is tagged in advance.

Modules used are Gazetteer, Query Analyzer; it supports SSO; rights are managed by RESTo itself, and has cart also for ordering. All the Open Source is on github.com.

Richard said this tool will be used for the RO. Also for the French Copernicus ground segment.

Jerome demonstrated how to create a collection, and post a resource in the collection. He concluded with:

* Shareable software
* Fast track for server implementation
* Fast track for client implementation

Yoshiyuki commented that some really good open source and reusable source has been exposed in this session that WGISS can post.

**Action WGISS-38-5**: WISP to organize a webpage for Open Source code, organized into categories (e.g. Interoperability). WGISS IG and Project leads to add links. These links should be the same as the links to Open Source code found on their webpages.

# WGISS Plenary Session, Part II

## WGISS Chair Summary

Richard Moreno provided the following summary of the meeting:

* Recovery Observatory
	+ Planning / roadmap
	+ Request for support from other agencies : data, software, and architecture
	+ Data cube could be an idea : track historical evolution
	+ CEOS CEO report
	+ 2 minutes video
	+ Adding of columns in action sheet
* WGISS term of reference
	+ 2 documents
		- Abbreviated Term of Reference
		- Governance document
	+ + 3Y workplan
* GEO
	+ Simplify the list of document IN02-C1
	+ Contribution to IN03 (GCI)? CWIC+FedEO+IDN
	+ GEO portal will change
		- no more clearinghouse
		- Knowledge broker for EO
		- GCI provider telecon is the best place where to promote CEOS OpenSearch, CWIC, IDN, FedEO
* Interoperability IG
	+ No lead for Interoperability Interest Group
	+ 1 global web page for Interop IG, but will link to dedicated pages (CWIC, FedEO, Opensearch, IDN ….)
	+ Will have documents and software in interest group pages, but also keep a global list (maybe in the CEOS tools web page)
* IDN:
	+ ISRO, now has 14 DIF (Metadata) entries in the IDN (with more to be inserted).
	+ Progress in the development of EUMETSAT DIF entries in the IDN (The entries are in review.
	+ ESA entries are in the progress of being developed.
	+ ROSCOSMOS and Geoscience Australia will start the process of IDN DIF generation.
	+ IDN CSW server has receive over 1.5 million total page views from users, during the past year (January 1, 2014 through August 31, 2014)
	+ Also, the IDN APIs services have received almost 750,000 total views, during the same time period.
	+ The main GCMD system holdings are made up of 53.44% of CEOS Metadata entries (non-NASA data set entries) + 32.39% (NASA entries) = total of 85.83% of the all entries.
* Disaster IG creation
	+ Decision postponed
* External Support Interest Group
* Taken by Interoperability Interest Group
* Water Portal
	+ Continues its operation at "http://waterportal.ceos.org"
	+ "OpenSearch two step search" will be implemented using internal servers (based on "GI-Cat"), instead of using IDN (for the 1st step) as proposed at the last meeting.
* OpenSearch
	+ We have several implementations, one best practice document. Then now, the 3 will align on the best practice
	+ First version CEOS Opensearch Best practice is issued for review. Will be completed and revised. At the end, there are examples of implementation : to be completed by the implementers + public period comment to global CEOS + to be tested each other to check that it works together
	+ Will add documents and available software in OpenSearch web page. A general « Interoperability IG » web page shall be done.
* FedEO
* Fedeo provides brokered discovery, access and ordering capability to European & Canadian EO mission data based on HMA standard interfaces. It also implements the OpenSearch OGC (and other) interfaces for an increased number of discoverable and accessible EO data collections, and for interfacing with CEOS Community Catalogues and Clients. More than 500 collections are now accessible through the system which has been also connected to the GEOSS Discovery and Access Broker for access in GEO. The system is being consolidated and is accessible for demos and administration through a dedicated Client Portal: <http://services-test.eoportal.org/web/guest/fedeo-demo-services>.
* Opensearch Protocol has been implemented (OGC 10-032r8 Geo and time Extension and OGC 13-026r4 Extension for Earth Observation) and provides currently access to HMA Catalogs supporting OGC 06-131 (Atom with EOP O&M (OGC 10-157) metadata as foreign markup or atom:link), CWIC catalogs (Atom with DC or ISO metadata as foreign markup or atom:link) and OpenSearch catalogs (e.g. Virtual Archive 4 and G-POD: Atom or RDF). The interface is aligned with OASIS searchRetrieve 1.0 conventions and additional RESTful interface is being prepared. Information on Opensearch implementation available at: <http://fedeo.esa.int/opensearch/readme.html> (Main), <http://geo.spacebel.be/opensearch/readme.html> (Test)
* FedEO Client guide and data partner guide available describing the steps needed to ensure accessibility to agency datasets through FedEO. Further catalogues interfacing is ongoing (e.g. DLR, EUMETSAT, VITO, NASA ECHO, Disaster Charter, and SPIRIT). Next planned activities will focus on the population of the dataset series catalogue with additional dataset series metadata (currently not all dataset series have ISO19139 metadata), on the integration of additional back-ends, on the improvement of the interfaces, on the integration with CNES RESTo Query Analyzer and Web browser search box and on the publication of FedEO relevant information and documentation on the CEOS WGISS Web Site.
* DSIG
	+ NASA has developed a Preservation Data Content Specification in Earth Observation
		- Introduce NASA data stewardship Documents in DSIG and in GEO
	+ LTDP documents to be reviewed by WGISS
		- Preservation workflow
		- Guideline
		- E2E consolidation

PID Best Practice Document

* + PURGE Alert
		- Depending on the case, the decision of purging could be taken by CEOS principles
* Emerging Technologies
	+ 2 real examples of big data in space : GAIA
	+ Data cube: not anymore at the image level, but at the pixel level. Opensource.
	+ SEO: spread data cube to make a network?
	+ Progressing well
	+ A new partner : EUMETSAT
	+ Significant progress has been made with ROSCOSMOS
	+ ISRO has connected MOSDAC to CWIC
	+ CWIC is an active partner of CEOS OpenSearch and support CEOS OpenSearch
	+ Big Earth data initiative NASA
		- Received funding from office of OSTP technology (White House)
		- GIBS as a opensource visualization tool
		- Promote use of OPeNDAP / [www.w10n.org](http://www.w10n.org/)
	+ White paper
		- For plenary : this is the next problem ; a lot of agencies are interested : this is the next key problem to be addressed by WGISS after data access and discovery
		- The objectives of CEOS will be met only if this challenge is met
		- 1 pager will be proposed to the side meeting participants
	+ WGISS summary
		- A new one will be issued in the frame of the ToR
* Agency Reports
	+ JAXA
		- ALOS-2 launch 24 May 2014
		- GPM Launched 28 Feb 2014 – data can be freely accessed from G-Portal
		- GSMaP : Global Satellite Mapping of Precipitation
	+ CAS / AOE
		- MODIS fusion : enhancement
		- ArcSer for response to Disasters: Earthquake on 4 August
	+ ESA
		- Status on Earth Explorer missions
		- Status on Copernicus : dissemination of S1 data : today
		- [http://scihub.esa.int](http://scihub.esa.int/)
	+ GA
		- 600 staff / ~150 AUD M$ ~ USGS
		- EO program! AUD 7M$ and 35 staff
		- Ground stations / 2 PB of EOS data
		- Open policy of data and technics
	+ NOAA
		- Thredds
		- Status of SAR constellation
		- Release of high resolution rapid refresh weather models
	+ NASA
		- GIBS & WorldView: opensource
		- NEW Earth Data Search Client : CMR+GIBS+OpenDAP and other services
		- NASA will be the US mirror of S1 S3 and S5p Copernicus data

**Action WGISS-38-16**: WGISS Exec to discuss the possibility of creating an interest group for Disasters.

## Adjourn

Richard thanked Roscosmos and Tamara Ganina for the wonderful hosting. He reiterated WGISS’ eagerness to continue cooperation with the Roscosmos team. WGISS continues to work toward wider cooperation, great and good activity, helpful for all.

## **WGISS-38 Actions**

Michelle Piepgrass reported that all actions from WGISS-37 are closed, and listed the actions resulting for WGISS-38.

|  |  |  |  |
| --- | --- | --- | --- |
| Action Number | Action Description | Actionee | Due Date |
| WGISS-38-1 | Richard to discuss with WGDisasters how the RO interacts with other disaster systems, not just at the agency level, but also UNSPIDER, Sentinel-Asia, Google (this is a strategic level question).  | Richard Moreno | 31-Oct |
| WGISS-38-2 | Martin and Andy to ensure a GoToMeeting is scheduled and publicized to WGISS-38 participants for the side meeting at the Plenary. | Martin Yapur, Andy Mitchell | 27-Oct |
| WGISS-38-3 | WGISS Exec to review the GEO task sheets to identify areas where WGISS can contribute.  | WGISS-Exec | 21-Nov |
| WGISS-38-4 | WISP to organize a teleconference to gather thoughts on best way to advertise important documents or recent document releases on CEOS website, and discuss with Kim Keith at the CEOS Plenary. | WISP | 30-Nov |
| WGISS-38-5 | WISP to organize a webpage for Open Source code, organized into categories (e.g. Interoperability). WGISS IG and Project leads to add links. These links should be the same as the links to Open Source code found on their webpages. | WISP | 21-Nov |
| WGISS-38-6 | WGISS Exec to identify a lead of the Interoperability Interest Group. | WGISS-Exec | 21-Nov |
| WGISS-38-7 | Lead of Interoperability Interest Group to develop a page on the WGISS website. | Lead of Interoperability Interest Group | 31-Dec |
| WGISS-38-8 | Interest groups and projects to update their pages on the WGISS website. | Interest groups and projects | 30-Nov |
| WGISS-38-9 | Michael Morahan to check whether the SKOS description of satellite and missions is in the GCMD and IDN. | Michael Morahan | 31-Oct |
| WGISS-38-10 | Andy to review the use of the term DOI in the GCMD DIF metadata field to determine if it is appropriate to use PI instead. | Andy Mitchell | WGISS-39 |
| WGISS-38-11 | WGISS to develop a glossary of terms. | WGISS-All | WGISS-39 |
| WGISS-38-12 | Andy write a summary of what NASA is doing with OPeNDAP (UNIDATA) and to organize a session on OPeNDAP at WGISS-39 (inviting presenters from OPeNDAP). | Andy Mitchell | WGISS-39 |
| WGISS-38-13a | Michelle, Ken, Jonathon and Wyn to finalize the “What is WGISS” text by end of October. | Michelle Piepgrass, Ken McDonald, Jonathon Ross, Wyn Cudlip | 27-Oct |
|
| WGISS-38-13b | Michelle, Ken, Jonathon and Wyn to complete ToR for submission to CEOS Plenary. | Michelle Piepgrass, Ken McDonald, Jonathon Ross, Wyn Cudlip | WGISS-39 |
| WGISS-38-14a | Andy and Jonathon to investigate an existing project for a few GFOI or GEOGLAM prototypes that would benefit from data exploitation that WGISS could promote. | Andy Mitchell, Jonathon Ross | Feb-15 |
| WGISS-38-14b | Technology Exploration IG to work with Brian to investigate options for supporting cloud processing, so Brian can present these to Barbara Ryan with a common voice between WGISS and SEO.  | Technology Exploration IG | Feb-15 |
| WGISS-38-15 | Jonathon to get the GA holdings into the IDN and available through CWIC. | Jonathon Ross |   |
| WGISS-38-16 | WGISS Exec to discuss the possibility of creating an interest group for Disasters. | WGISS Exec | 27-Oct |
| WGISS-38-17 | WGISS-All to provide comments on the draft OpenSearch Best Practice document by end of November | WGISS-All | 30-Nov |
| WGISS-38-18 | WGISS-All to provide comments to Preservation Workflow once they are sent out by Mirko. | WGISS-All | One month after delivery of document. |
| WGISS-38-19 | DSIG to finalize the purge alert process using CEOS principles. | DSIG | WGISS 39 |

# Glossary of Acronyms

API Application Programming Interface

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DIF Directory Interchange Format

ECV Essential Climate Variable

EO Earth Observation

ESIP Federation of Earth Science Information Partners

GCI GEOSS Common Infrastructure

GENESI Ground European Network for Earth Science Interoperations

GEO Group on Earth Observations

GEO-GLAM Global Agricultural Monitoring

GEOSS Global Earth Observation System of Systems

GIS Geospatial Information System

GPM Global Precipitation Mission

GSDI Global Spatial Data Infrastructure

GUI Graphical User Interface

IDN International Directory Network

IG Interest Group

ISO International Standards Organisation

LSI Land Surface Imaging

OGC Open Geospatial Consortium

PoC Point of Contact

SEO Systems Engineering Office

SBA Societal Benefit Area

SDCG Space Data Coordination Group

SIT Strategic Implementation Team

SST Sea Surface Temperature

ToR Terms of Reference

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