MINUTES

OF THE

39th MEETING

OF THE

CEOS WORKING GROUP ON   
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Tsukuba, Japan

11 May to 15 May, 2015

Hosted by

Japan Aerospace Exploration Agency

(JAXA)

Table of Contents

[1 WGISS Plenary Session, Part I 5](#_Toc426318860)

[1.1 WGISS Welcome, Introductions, Adoption of Agenda 5](#_Toc426318861)

[1.2 Logistics Information 5](#_Toc426318862)

[1.3 Host Opening Address 5](#_Toc426318863)

[1.4 WGISS Infrastructure Support Project (WISP) 5](#_Toc426318864)

[1.5 WGISS Chair Report 5](#_Toc426318865)

[1.6 CEOS Executive Officer (CEO) Report 6](#_Toc426318866)

[1.7 Systems Engineering Office (SEO) Report 6](#_Toc426318867)

[1.8 GEO Task Sheet Review 7](#_Toc426318868)

[1.9 GEO Work Programme 7](#_Toc426318869)

[1.10 GEO Secretariat Report 8](#_Toc426318870)

[1.11 WGISS Organization Discussion 8](#_Toc426318871)

[1.12 WGISS Website, CEOS Data Management System 9](#_Toc426318872)

[1.13 WGISS Terms of Reference 9](#_Toc426318873)

[1.14 Future Meetings 9](#_Toc426318874)

[2 Agency and Liaison Reports 10](#_Toc426318875)

[2.1 Japan Aerospace Exploration Agency (JAXA) 10](#_Toc426318876)

[2.2 Russian Federal Space Agency (ROSCOSMOS) 10](#_Toc426318877)

[2.3 National Aeronautics and Space Administration (NASA) 10](#_Toc426318878)

[2.4 Centre National d’Etudes Spatiales (CNES) 11](#_Toc426318879)

[2.5 European Space Agency (ESA) 11](#_Toc426318880)

[2.6 UK Space Agency (UKSA) 11](#_Toc426318881)

[2.7 Global Spatial Data Infrastructure (GSDI) Association 12](#_Toc426318882)

[3 WGISS Projects 12](#_Toc426318883)

[3.1 CEOS Water Portal Project 13](#_Toc426318884)

[3.1.1 Operation Status 13](#_Toc426318885)

[3.1.2 Demonstration 13](#_Toc426318886)

[3.1.3 Discussion 13](#_Toc426318887)

[3.2 Disasters Recovery Observatory (RO) Project 13](#_Toc426318888)

[3.2.1 Main Features 13](#_Toc426318889)

[3.2.2 Technologies and COTS 14](#_Toc426318890)

[3.2.3 Technical issues 14](#_Toc426318891)

[3.2.4 Future Features 14](#_Toc426318892)

[3.2.5 Live Demonstration 14](#_Toc426318893)

[4 Interest Groups 15](#_Toc426318894)

[4.1 Technology Exploration Interest Group 15](#_Toc426318895)

[4.1.1 Cloud / Big Data Agency Update 15](#_Toc426318896)

[4.1.2 Data Management Plan 16](#_Toc426318897)

[4.1.3 Open Source Software (OSS) 18](#_Toc426318898)

[4.1.4 OpenDAP 19](#_Toc426318899)

[4.1.5 Data Cube Workshop 19](#_Toc426318900)

[4.2 Data Stewardship Interest Group 22](#_Toc426318901)

[4.2.1 Overview and Updates 22](#_Toc426318902)

[4.2.2 Purge Alert Procedure 23](#_Toc426318903)

[4.2.3 CEOS Best Practices 23](#_Toc426318904)

[4.2.4 Persistent Identifiers Pilot Implementation 24](#_Toc426318905)

[4.2.5 Preservation of Software and Documents at CEOS Agencies: Approaches and Lessons Learned 26](#_Toc426318906)

[4.2.6 ESA Approaches to Preservation 26](#_Toc426318907)

[4.3 CEOS Interoperability Interest Group 27](#_Toc426318908)

[4.3.1 International Directory Network 27](#_Toc426318909)

[4.3.2 CEOS WGISS Integrated Catalog (CWIC) 28](#_Toc426318910)

[4.3.3 Federated Earth Observation (FedEO) 32](#_Toc426318911)

[4.3.4 CEOS OpenSearch Project 32](#_Toc426318912)

[5 WGISS Plenary Session, Part II 33](#_Toc426318913)

[5.1 WGISS Chair Summary **Error! Bookmark not defined.**](#_Toc426318914)

[5.2 Adjourn 33](#_Toc426318915)

[5.3 WGISS-39 Actions 33](#_Toc426318916)

[6 Glossary of Acronyms 34](#_Toc426318917)

List of Participants

CAS-AOE Chaoliang Wang, Guangyu Liu

CEOS-DCEO Jonathon Ross\*

CEOS-SEO Brian Killough\*

CNES Richard Moreno (WGISS Chair), Jérôme Gasperi, Mathieu Gond, Michelle Piepgrass (WGISS Secretariat)

DLR Tyler Christensen\*

ESA Mirko Albani, Yves Coene\*, Philippe Mougnaud\*

GEO Secretariat Osamu Ochiai\*

Geoscience Australia Simon Oliver\*

GSDI/HUNAGI Gábor Remetey-Fülöpp

INPE Lubia Vinhas\*

ISRO Nitant Dube\*

JAXA Satoko Miura, Ichiro Naito, Yoshiyuki Kudo, Shinichi Sekioka, Kaori Kuroiwa, Akari Yoneyama

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

Michael Morahan, Doug Newman\*, Archie Warnock\*, David Fulker\*, James Gallagher\*, Liping Di\*

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

Russian Space Systems Tamara Ganina

UKSA Wyn Cudlip

USGS Kristi Kline\*

\* Via web conference or email

# WGISS Plenary Session, Part I

## Welcome, Introductions, Adoption of Agenda

Richard Moreno (WGISS chair) welcomed the participants to WGISS-39. Richard thanked Satoko Miura for all the excellent arrangements for the meeting, and asked those present to introduce themselves. He reviewed the agenda and it was adopted with no significant modifications.

## Logistics Information

Satoko Miura welcomed the WGISS-39 participants to the JAXA Tsukuba Space Center and described the logistics of the meeting for lunch, breaks, group photo, dinners and evening activities. She also made a few suggestions for sightseeing and provided emergency contact information. Satoko noted an unseasonal typhoon that is forecasted to pass by the Tokyo area on Wednesday.

## Host Opening Address

Mr. Ichiro Naito, Director of Space Applications and Operations Center (SAOC), JAXA, welcomed the participants on behalf of JAXA. He stated that the main roles of SAOC are to promote satellite data utilization supporting public/private sectors efforts to solve issues by using satellite data; to develop and operate systems for disaster prevention/mitigation, including emergency operation for major domestic/foreign disasters; international coordination and cooperation with Asian countries; to develop, operate and maintain Earth observation satellite ground systems, including data processing, data storage and data distribution. SAOC also has the role of coordination with external bodies, evaluation of new technologies and technological trends related to the above.

Mr. Naito explained that JAXA experienced a significant reorganization on April 1. He displayed a diagram showing the main points, and the organizational structure with Mr. Shizuo Yamamoto as Vice President and Director General.

Mr. Naito listed JAXA’s in-orbit and future Earth Observation missions. He gave examples of observations from some of these missions:

* Time lapse showing the shrinking of Arctic sea ice concentration at an unprecedented rate from GCOM-W;
* GSMaP (Global Satellite Mapping of Precipitation) images of hurricanes Ingrid and Manuel;
* GCOM-C, a new mission with an SGLI (Second-generation Global Imager) to observe aerosols and clouds, as well as vegetation and temperatures in the land and ocean region;
* SGLI-derived phytoplankton, aerosol and vegetation activity that will be used for mapping fisheries, monitoring crop growth and estimating crop yield;
* EarthCARE/CPR in cooperation with ESA to reduce the uncertainties in global warming prediction by measuring the three dimensional structure of clouds and aerosols, which are most uncertain parameter in the numerical climate models;
* GOSAT (Greenhouse Gases Observing Satellite), with a Thermal and Near Infrared Sensor (TANSO) for carbon Observation;

Mr. Ichiro Naito thanked everyone for their attention, and wished the participants for a fruitful meeting.

## WGISS Infrastructure Support Project (WISP)

Martin Yapur gave the report of the WGISS Infrastructure Support Project. He listed the members of the team, and gave the link to the wiki where presentations can be loaded: <https://wiki.earthdata.nasa.gov/display/W3M/WGISS+39+Meeting+Home>. The presentations will be migrated to the Google drive for viewing/archiving after the conference; they will also be accessible from the agenda on the WGISS page.

The project’s current activities include updates of existing webpages to coordinate with the new look of the CEOS website, creation of new pages for FedEO project and Open Source Code, organization/upload of documents to the new Document Management System for archival, continual updates of the mailing lists, and support to the CEOS Document Management System.

WISP seeks WGISS support to continue managing the main content of the WGISS website and the mailing lists, and to generate and support outreach materials.

Martin noted continued technical problems with the WGISS email distribution lists that they are working to resolve. Andy asked if it is something that can run as a CEOS service that the SEO should be offering to all the CEOS WGs.

**Action WGISS-39-01**: WISP to talk to SEO about transferring the WGISS email distribution lists to CEOS and the WGISS documents to the CEOS Document Management System.

## WGISS Chair Report

Richard Moreno gave the WGISS Chair report, outlining the current work of WGISS, which should be concrete and have visible output, and useful and an asset to WGISS activities. WGISS should strive to valorize its work through the CEOS portal.

Richard reported the following successes for WGISS:

* Opensearch Best Practices document is ready to be issued, and is aligned with the various OpenSearch implementations. A guide for developers is forthcoming.
* Data Stewardship Interest Group provides concrete output from LTDP; three documents for inclusion as best practices have been produced.
* CWIC and FedEO are two examples of concrete support provided to CEOS agencies.
* Recovery Observatory Infrastructure is an effort in support of WGDisasters, and already has concrete output.
* Emerging technologies affect every agency; WGISS presentations and discussion allow WGISS to understand what others are doing, and help avoid duplication of effort.
* In terms of outreach in 2014, a side meeting around the plenary was conducted, along with a white paper.
* IDN is an operational system, and the most important contributor to GEO. It has undergone strong evolutions in 2014/2015.
* Water portal is operational.
* WGISS opensource web page has been set up, and WGISS is actively working to add elements.
* Continued support to the other CEOS bodies: WGs, VCs. The SIT team has the objective to increase the synergy among WGs and VCs. A VC/WG side meeting was held in 2014, and another SIT side meeting is foreseen in September 2015. Andy added that he is developing a ToR on a liaison from WGISS to these groups. WGDisaster relationship is good due to participation in the Recovery Observatory. The VCs do not express a clear need for support. WGCapD will be assisted by the opensource web page. WGCV relationship will be improved with a joint meeting in 2015, WGISS-41.
* DataCube provides a wonderful perspective: work at the pixel level, not only at the file level. This is an occasion to really enter the era of big data and how can it be useful at the agency level. WGISS is presented with two different approaches: SEO with a specific activity linked to Kenya DataCube, and the GA DataCube activity.

Wyn Cudlip commented that GEO is not listed much; Andy Mitchell added that there are support activities to GEO that are not listed, and it would be good to add those. Richard agreed that WGISS should outline them more. Mirko Albani said he was at the symposium in Geneva, where there is continued work on the GEO Work Plan, and WGISS is in a position to help shape the CEOS contribution. Richard asked Mirko to upload his presentation to the wiki, and focus a little more on GEO during his presentation.

## Deputy CEOS Executive Officer (DCEO) Report

Jonathon Ross gave the Deputy CEOS Executive Officer report, discussing what is happening in the broader CEOS world. He began saying that 2016 will be a big year. CEOS now has the GEO 2015-17 Work Plan; there is a huge amount of work there for CEOS and for WGISS as well. The carbon section covers 42 actions for CEOS. Many years went into developing CEOS carbon strategy and if a WG thinks they cannot take on an action it is the working group’s responsibility to notify CEOS. For the GEO strategic plan, CEOS provided a strong response saying it will continue to be the space arm in support of GEO. How the GEO work program will work is still unknown, but there was a symposium in Geneva where it was clarified how CEOS will work in the new world of GEO. Jonathon mentioned that CEOS now has a new ‘deliverables online tracker’. Chair and vice chair of working groups can log in and provide updates; one update is due in May.

Jonathon urged WGISS to work with SEO to keep the content of the web site up to date. There are new templates and logos to use, and a new CEOS brand. He also noted that all the CEOS working groups need to have a ToR following the CEOS template by the end of the year.

Andy asked if the GEO 2016 Work Plan is now stable. Jonathon replied that a lot of the structure and terminology will change, but there are not likely to be changes in the work and activities. Mirko asked if he plans to collect inputs from WGs and then send the CEOS response. Jonathon replied that he has not had a chance to talk to Marie Jose about this. The review process in CEOS will commence soon, and his understanding is that WGISS should expect a new draft, and from that there will be a discussion across CEOS.

## Systems Engineering Office (SEO) Report

Brian Killough, Systems Engineering Office, gave the following report of the activity of the Systems Engineering Office. Keeping a steady count of the CEOS missions, CEOS currently operates 131 missions, had 12 successful CEOS launches in 2014, and 12 new CEOS launches are planned for 2015.

COVE, now with 259 missions, 697 instruments, and 20,000 unique users in 2014, has new features: Data Overlays of Landsat tiles, Sentinel-2 tiles, ASTER DEM, GlobCover, MODIS LCC, future overlays of vegetation phenology, and cloud probability maps for agriculture. Data archive links: Landsat 7/8, SPOT 1-6, Pleaides-1A/1B, Radarsat-2, ALOS-1 (thanking WGISS for helping to make these connections possible). Future data archive links are Envisat, ERS-1/2, RapidEye, and Sentinel-1A/2A.

TheCoverage Analyzer Tool, a needed capability to view individual countries, has a new release available with an improved menu selections for region, past/future, dates, missions, discretization, improved processing performance, output in global or table format, selection of cloud filtering, color scaling, and KML and CSV export.

The Data Policy Portal now has columns with links to client portals, EO portal, IDN, etc. A new release of the Data Policy Portal is available. Andy asked if it only maps to CEOS missions, and if so, how is a CEOS mission defined. He added that a CEOS mission can be one launched or organized by a CEOS agency, but sometimes they will partner with a CEOS agency. It was agreed that if a CEOS agency is involved, then it is a CEOS mission. Brian noted that the portal only relates to mission/instrument combinations, not actual datasets, and it only applies to satellite data (active or passive). The portal reports Open Access = 59%, Restricted or Unknown = 41%. The Database includes 352 missions launched since 1990 and 646 mission-instrument combinations. The tool includes direct links to primary data portals.

Brian also reported on the FAO Project, where the stakeholders wanted an easier access to specific data, and approached CEOS to see if this could be done. The prototype has a traditional scene-based file manager and user interface, user and group control, terminal line editor for OpenForis toolset, and search tool for Landsat archive searches and data download. Initial delivery was in October 2014 and deployment to four countries in early 2015. Expanded deployment to 14 countries is planned through 2018, led by FAO.

For the Asia-Rice Project, SEO developed a data services portal with secure access for JAXA and Indonesia Asia-RICE science users, and dedicated storage for up to 35 Radarsat-2 images (provided by CSA) and Amazon cloud processing of SAR datasets using INAHOR tool. Datasets and processing tools are installed, but initial tests in Indonesia showed poor processing performance due to internet and graphics demands of the INAHOR processing tool. They moved the server to Japan in late 2014, and are now considering moving the server to Singapore for improved performance. The Asia-RICE data services portal will be expanded in 2015 to include Vietnam, India, Indonesia, Japan, Lao PDR, Malaysia, Philippines, Taiwan and Thailand. There will be separate data folders for each country.

The SEO is also working on the Colombia Project. The SEO developed a proposal to DLR for TanDEM-X datasets over Colombia in March 2014; it was accepted in June and in December they delivered the data services tool to Colombia including downloaded datasets. They are using Amazon cloud storage and processing with secure access to datasets in eight regions over Colombia. CEOS facilitated data access and Amazon Web Services to demonstrate enhanced computing performance and improved digital elevation model data. Feedback is expected in mid-2015.

JECAM Project: The SEO worked with CSA to develop a data sharing agreement for Radarsat-2 data. The approved document will allow sharing of datasets across all JECAM sites. In the past, data was restricted to only individual sites based on separate proposals. The SEO plans to develop a user interface for file storage and data processing to support a SAR inter-comparison study for JECAM. The data services tool will support Canada, Ukraine, Belgium, Argentina, China and France. There will be separate data folders for each country. This provides a broader group of users to support JECAM.

Andy asserted that a session on open source and hurdles agencies have come across with sharing source is planned, and wondered if Brian has any lessons learned. Brian replied that he has had a number of discussions with NASA on how this can best be done; in the DataCube project they knew that Australia had already found a mechanism. Brian found two things that worked: Apache v2.0, and General Public Release. This is even more open than Open Source, and a good thing to target since it is wide open. NASA Open Source is very burdensome. Andy said that he found that every center has different rules. Brian suggested talking off-line about this. It is a tough process but if some of these hurdles can be solved everyone will be better off. Brian will send information on Open Source to WGISS.

## GEO Task Sheet Review

Richard Moreno led a presentation and discussion on the GEO Task Sheet. He noted that the GEO Secretariat is requesting an update on the progress and the achievements associated with the GEO task components for which WGISS acts as Point of Contact, as well as plans and perspectives for the future, adding that 2015 will be a year of transition from the first to the second GEO decade. While renewing the GEO mandate for the period 2016-2025, the Geneva 2014 Summit asked that a new ten-year plan be developed.

Richard began with a diagram showing the 2012-15 GEO Work Plan management structure. Richard Moreno is listed as the PoC of the component IN-02-C1. He listed the key outputs, and the challenges hindering progress. He noted that they have started to establish Data Management Principles based mainly on IN-02-C1. He proposed going through the Earth Datasets Component (C1) sheet (<http://earthobservations.org/ts.php?id=135>) which shows milestones and status of key outputs and key activities.

The question was raised that some of the outputs are documents that have been generated but it is not known how to distribute these to wider community, and how to replace documents with later versions. GEO and CEOS are working on that. There are a dozen documents for data preservation, and only two should be enough. One possible option may rest with the data management principles that will be approved by GEO. It may be a good thing to link or point to more specific technical procedures within the overarching document.

Richard said WGISS can propose synthesizing some of the list, and Mirko wondered if it is better to begin focusing on the 2016 Work Plan, and these can be the materials to feed for 2016. Wyn asked if it is possible to just provide access to these documents, linking them to the component sheet, and start a new sheet for those relevant to 2016. Satoko suggested two systems for the document management: one for management, and one for the users. It was noted that there are some points on quality, but in the new work plan, quality is separated out.

Andy recommended that a small group of WGISS people completely redo and make recommendations of what should be done for 2016, preparing the list for a smooth transition to the 2016 Work Program. GEO should be the repository for these documents. GEO should provide the resources for the document repository.

**Action WGISS-39-02**: Mirko Albani, Richard Moreno, Andy Mitchell and Wyn Cudlip to select a team that will review GEO Task IN-02-C1: Advances in Life-cycle Data Management, and make recommendations looking toward the GEO 2016 Work Plan.

## GEO Work Programme

Mirko Albani gave a presentation of the GEO Strategic Plan 2016-2025: Implementing GEOSS and GEO Work Programme 2016. This new plan was developed with the following characteristics:

* Unique value: Improve the quality, timeliness, range and availability of Earth observations, facilitate the development of efficient, sustainable solutions to environmental societal challenges, and broker combinations of actors to understand Earth system processes and respond to societal challenges.
* Strategic objectives: Three strategic objectives in the areas of action – Advocate, Engage, Deliver; Encompass major GEO achievements and success.
* GEOSS: In its role is Observation and Information Systems, GEO implements central components, based on clear high-level principles, and name principles for central GEOSS components: openness, effectiveness, flexibility, adaptability, sustainability, reliability.
* Advocate: With core function for data sharing, data management, gap analyses, best practices for citizen science, radio frequency protection.
* Engage: With core functions for creating partnerships, resource mobilization, creating visibility and awareness, broadening the user bases, capacity building, and promoting EO use in development activities.
* Deliver: With core functions for operating common elements of GEOSS, documenting data quality and provenance, maintaining a knowledge base, conducting implementation pilots, incubating services through initiatives.
* Implementation mechanisms: Defined four implementation mechanisms, GEO Community Activities, GEO Initiatives, GEO Flagships, and GEO Foundational Tasks.
* Revised SBA, since current SBA did not fully enable coordination and user-engagement.
* Work Programme into 3-year Work Programmes, transitional Work Programme for 2016, that includes all GEO actions, decide on priorities, discuss and agree in GEO Programme Board, and include indicative resource planning.
* Governance: Needs to clearly support implementation needs.
* Resourcing: Commitments maintained as voluntary, support for use of an indicative scale of contributions, ideally multi-year, commitments to be specified for flagships and initiatives, resource allocations specified in Work Programme

Mirko next described the GEO Work Programme, which consists of foundational tasks, flagships, initiatives, community activities, and resource summary. He described each of these, and gave a few examples. He also described the key differences in terms of purpose, originator, steering and oversight, and reporting.

Mirko also described the process, leading up to adoption at the GEO Plenary. Work Programmes will generally span three years. He also compared the GEO Work Plan to the GEO Work Prgramme, and highlighted what is new in the programme.

Mirko gave some elements of the current draft 2016 Work Programme, which has been compiled by the GEO Secretariat and has been drafted to comply with the structure suggested by Annex B of the draft “GEO Strategic Plan 2016-2025: Implementing GEOSS”, issued by the IPWG in March 2015. The content of the Plan is based on the inputs received by the current GEO Tasks Team (provided through a template that the Secretariat circulated in mid-March and complemented by the Secretariat), and on a preliminary analysis of the Foundational tasks necessary to implement all the GEO Core Functions identified in the new Strategic Plan, performed by the Secretariat. He listed candidate foundational tasks, community activities, candidate GEO Initiatives, and candidate GEO Flagships

Andy will need to ask Jonathon if WGISS is the appropriate group to lead Foundational Task 2 (IN-02-C1), and determine how WGISS fits into it. It is important that the task description contain tasks where WGISS can really contribute, and that it list concrete things WGISS can support.

## GEO Secretariat Report

Osamu Ochiai gave an update on GEO, including the state of the GEOSS infrastructure. He displayed a diagram showing the enabling of the System of Systems, where new elements include the GEOSS Clearinghouse, IDN, and CWIC. GEOSS assets include about 38 brokered data providers, more than 41 million accessible resources (data collections and datasets) which contain more than 174 million assets. He noted that it will be important to demonstrate the accuracy of these numbers. They have encountered difficulties tracking the source of large reported numbers in the past and verifying that they were actually accessible. He showed a graph of GEOSS resources, showing top contributions from FedEO and CWIC.

The GEOSS Portal has undergone major functional and layout improvements of main components. Data discovery is through collections and granules, and data access preview and download can be done as is, or as a transformation. GeoRSS feed is displayed on the map's home page, there is contextual help and user registration (using existing social credential or log in with a newly created account).

The GEOSS Portal allowsusersto select a set of data from different data sources; via the web service exposed by the GEO DAB (subset, format conversion, CRS transformation and data interpolation) the user is able to define a common grid*.* The selected data are downloaded according to the defined common grid.

Osamu also listed a number of new standards/protocols now supported by the DAB, such as ESRI Map Server, Grads Data Server, CKAN, DCAT, BCO-DMO, IRIS Services (Event, Station), Hydro Server/HIS Central profiler, New GBIF REST API, HTTP-GET implementation of CSW/ISO, and Generic Web Accessible Folder for shape files.

Data providers under coordination are FP7 GEOCARBON, IGN, FAO, UNEPLive, China, UK.GOV, and GOSIC. Brokered providers are WMO/WIS, US Data.Gov, and DIAS/Japan. Private Sector providers are Pre-Ops: ESRI ArcGIS Users, Digital Globe, e-GEOS CSM, Microsoft, StormCenter, and Blackbridge. Potential providers in future are Copernicus, India, Russia, and AfriGEOSS.

The criteria for prioritization of data providers to be brokered in 2015 is key/targeted datasets from SBA community (identification and engagement), big and government data providers including private sector, and data providers willing to contribute to, and to get benefit from, GEOSS.

Five data management principles were listed and defined. DMPTF will prepare implementation strategies for discussion at GEO-XII Plenary.

Audience feedback on the GEO Workplan provided for sustaining existing GEOSS Services and to support the evolving 2025 GEOSS vision and infrastructure. Key messages are that understanding of stakeholder groups is needed, as is the delivery of high end products (indicators, information), improvement of communication and participation, and SBA link and integration to the infrastructure team. In terms of data, need to address key data services (Data Access and Use, Data Integration, and Data Reliability). From data to information products, additional new sources of data are needed. In terms of infrastructure, the system of systems should build on existing but evolve to meet new needs, should support real-time collection, and community portal alignment is needed.

Osamu listed the priorities and for the 2016 Work Programme, noting that there should be demonstrable value of GEOSS for GEO community and beyond in order to be successful.

Osamu added that they welcome contributions from more space agencies, welcome implementation on technology evolution (Processing, Integration, Analysis, Knowledge base, etc.) with use cases (community services/portals), and welcome inputs to the post 2015 GEOSS Vision and Architecture. These are areas where WGISS can contribute.

Andy asked about the source of the metrics on GEOSS resources showing CEOS WGISS and FedEO; Osamu replied that they are from DAB metrics. Andy also asked Osamu if he has any advice for WGISS as it develops inputs to the GEO 2016 Work Programme. Osamu noted that GEO needs to enhance the interaction with the users, with use cases and community portals. The focus has been discovery and access; it is time to go beyond that, and community services is a good perspective to work toward.

Mirko commented that WGISS reviewed the foundation tasks, and asked Osamu to explain the transition from IN-02-C1 to FT2. Osamu will be submitting a counter-proposal; 2016 is a transitional time. The Data Management Task Force is very active. IN-02-C1 tasks are a mixture, and the best framework needs to be identified to get synergy. Osamu also noted that it is too early to know who will be on the program board that oversees the FTs.

## WGISS Organization Discussion

Andy Mitchell led a discussion on discussion on changes to the organization of WGISS.

The Interoperability Interest Group needs a lead, and Andy suggested that the vice chair take the role. This person will coordinate the activities of the various projects (IDN, FedEO, CWIC, Water Portal, and OpenSearch). At WGISS-38 it was thought that this person could liaise with various groups in CEOS. Instead, WGISS-Exec decided that it would be better to have a WGISS Interoperability Liaison, and for that purpose Andy has developed a Terms of Reference. WGISS thought that Jonathon Ross would be the ideal person as he goes to all the CEOS meetings. Wyn suggested using a word different from Interoperability to avoid confusion, and suggested collaboration facilitator or liaison. The term WGISS Collaboration Facilitator was adopted. Agreement needs to be obtained from Jonathon, and it was suggested that CEOS SEC formally endorse it. There should always be a report from Collaboration Facilitator at WGISS meetings.

Richard raised the topic of keeping an interest group inside WGISS related to disaster. Karen Moe is retiring, and may not be replaced with a person who could join WGISS. Wyn suggested that the agencies see that there is a WGDisasters, so would not want to send someone to WGISS also for disaster work. Wyn noted that this does not rule out having an agenda item for Disaster activities.

**Action WGISS-39-03**: Richard Moreno and Andy Mitchell to approach Jonathon Ross about taking on the role of WGISS Collaboration Facilitator, then approach the SIT and the SEC for adoption.

**Action WGISS-39-04**: Michelle Piepgrass to include a session at WGISS-40 for deciding the future of the OpenSearch Project.

## WGISS Website, CEOS Data Management System

A discussion on how to improve the WGISS website, document management system, and email distribution lists resulted in the following points:

WGIS Website:

1. The short paragraph developed to introduce readers to WGISS is no longer there. It is still on old.ceos.org.
2. Old.ceos.org needs a disclaimer that this page has been replaced.
3. Home page should have brief information on the previous and upcoming WGISS meetings only.
4. Remove distinction between interest groups and projects from the website. WGISS will keep the distinction but it is not necessary on the website. Each should have a page, and each page should have:
   1. Purpose
   2. Scope
   3. Background
   4. Services
   5. Communication and Coordination
5. Need a page/link referencing old groups/projects work (completed/closed activities)

Document Management System

1. Review the DMS page – it implies you have no access to any documents when in fact you can upload if you click on a link to a document in the DMS.
2. Organize the documents according to the five GEO Data Management Principles:
   1. Discoverability
   2. Accessibility
   3. Usability
   4. Preservation
   5. Curation
3. Verify that old documents into the DMS.

Email distribution lists: Align with CEOS methodology

**Action WGISS-39-05**: Michelle Piepgrass to work with WISP to implement the recommended changes to the WGISS website, including the short WGISS “purpose/description” paragraph.

**Action WGISS-39-06**: IG/P leads to implement recommended changes to their pages on the WGISS website.

**Action WGISS-39-07**: Michelle Piepgrass to work with WISP to migrate WGISS documents to the CEOS document management system; also discuss with the SEO the wording on the CEOS documents landing page.

## WGISS Terms of Reference

Michelle Piepgrass agree to modify the WGISS ToR so that it follows the CEOS template. Several points were raised: change word “archive” to “preservation”, Remove “description”, simplify interest group and project creation procedures, and align the document with the five Data Management Principles.

## Future Meetings

Wyn Cudlip presented information on WGISS-40, to be hosted by UKSA, September 28 to October 2, 2015, in Harwell, Oxfordshire, UK, which is 40 km south of Oxford. The nearest airport is Heathrow, and taxi can be prearranged through the Catapult offices for better pricing; there are also trains. The meeting will be held at the Satellite Applications Catapult, a campus of 3000 employees. Participants will be picked up at the selected hotel (to be determined) and taken to the campus.

Andy Mitchell reported that WGISS-41will be a joint meeting with WGCV, March 14 to March 18, 2016 hosted by CSIRO/GA team Australia.  If the Australian Bureau of Meteorology becomes a CEOS associate this year they will be invited to support.

It was suggested that WGISS-42 be held in the Americas, and that a meeting with a Copernicus workshop might help bring in DLR.

# Agency and Liaison Reports

.

## Japan Aerospace Exploration Agency (JAXA)

Akari Yoneyama gave the Japan Aerospace Exploration Agency (JAXA) agency report. She introduced herself as the newest member of the JAXA team for WGISS.

G-Portal is the JAXA satellite data dissemination system. It features cross-search from JAXA’s many satellites, and provides many of JAXA’s satellite data free of charge. HTTPS and SFTP services are available. Akari listed the products provided by the G-Portal. The G-Portal has a support desk ([z-gportal-support@jaxa.jp](mailto:z-gportal-support@jaxa.jp)).

CATS-I is the catalogue transfer service interface of G-Portal. It provides the catalogue with thumbnails and browse. Supported interfaces are OpenSearch, ebRIMEO and ISO19115. CATS-I provides the catalogs of GPM, GSMaP, TRMM, and ALOS. Only specified users can search and obtain the catalog.

Akari also mentioned the JAXA reorganization, showing a diagram of the organization, and the organizational structure. JAXA WGISS members belong to Satellite Applications and Operations Center (SAOC).

Akari reported that PR/TRMM completed every mission, and finished observations on April 1, 2015. JAXA completed normal mission operations on October 7, 2014. After that, JAXA had conducted extra experimental operations. In the near future, JAXA plans to start the reprocessing and release of version 8 products via the G-Portal.

This year’s data release news: All of the GPM products including the past data will be updated to version 4. All of JERS-1 products will be reprocessed and provided via the G-Portal.

HAYABUSA-2 was launched May 10, 2014 and will plan for an Earth swing-by and start the exploration of deep space. ASTRO-H will be launched in 2015. Akari displayed a diagram of the long term plan of JAXA observation.

Yonsook asked if the access is part of the new data policy. Akari said yes, though the high resolution is still restricted.

## Russian Federal Space Agency (ROSCOSMOS)

Tamara Ganina gave the report for the Russian Federal Agency. She began sharing the latest news in the Russian space industry: At the beginning of the year the Russian space industry reconstruction process started. Now most of government and commercial organizations including JSC “Russian Space Systems” were included in United Rocket and Space Corporation. The formation of this system will be finalized to 2020. Igor Komarovwas appointed Head of Russian Federal Agency (Roscosmos). In March the government approved the new version of the law on space activities. This document confirms free access for all users to Russian satellite data with 30+ meters resolution. Commercial distribution of high resolution data for all customers is also possible.

Russian EO missions launched in 2014 were Meteor-M2, Resurs-P 2. Russian EO missions launched in 2015 are Resurs-P 3, Kanopus-V 2, and Electro-L 2. Russian EO missions to be launched in 2016 are: Meteor-M 2-1, Meteor-M 2-2, and Electro-L 3. Tamara displayed a table showing information capabilities of operational Russian satellite constellation.

Roscosmos GeoServices has been providing the Roscosmos Geoportal since 2010. From 2015, basic EO provision is from Data Products Bank, and the Open EO Data Portal is in process. Tamara displayed a diagram of significant international cooperation in the categories of Operational services, experience exchange, research projects, and data distribution.

Tamara also reported on the Earthquake Monitoring Project. This innovation project using satellite data examines some problems with prediction of earthquakes. The methodology uses the concept of seismotectogenesis (Doda L.N.), which is a combination of geophysical patterns and corresponding methods for calculating the parameters of possible earthquakes: the date, location and magnitude. Particular attention is paid to clouds and the magnetic characteristics of a possible earthquake. She described the geophysical patterns observed, noting that the connection of the magnitude of earthquakes and the maximum length of seismic clouds has been discovered near activated seismic location (predict intervals of magnitude of earthquake). At the activated section of the plate boundaries, blocks or faults under certain conditions begin to develop specific clouds that have the same contours of such boundaries and visible on satellite images in a variety of structures. In the example given, for the period of a few hours, the clouds are fixed in the places of occurrence and are not involved in the process of moving air masses.

She listed the main components of GIS for earthquake monitoring, and displayed the architecture diagram of the system. She listed the source of the gravimetric measurements, the proton measurements, the electro-telluric measurements, the space images, the measurements of Earth rotation parameters, and the geophysical supervision over the Sun. A thematic structure diagram was displayed, as was the portal that is used. Tamara indicated that this method requires careful manual analysis, and only have two specialists to work on this.

Mirko asked how many earthquakes were predicted and not confirmed; Tamara noted that they analysed only two areas to test the method; eight predictions were made, and eight confirmed. They have been looking at new test sites since September 2013. Perhaps at the next meeting Tamara can present results. The team has discussed internally about connecting with WGDisasters to give this method a broader application.

## National Aeronautics and Space Administration (NASA)

Andy Mitchell gave the National Aeronautics and Space Administration (NASA) agency report. Andy defined the role of the Earth Observing System Data and Information System (EOSDIS), and explained the structure. He noted that their data centers are discipline-oriented, and there are a number of science investigator-led processing systems distributed around the country. EOSDIS data products are processed at various levels ranging from Level 0 to Level 4, and products are formatted for use by the discipline community. Use of standard formats is required by NASA, but multiple formats still present issues to usage/interoperability

Data volume and velocity of data ingest and distribution are increasing rapidly. Process and infrastructure improvement are continuous. Andy described the EOSDIS user registration system which is being implemented bydata centers, subsystems and related services. He also displayed an outline of their “Vision 2020” for discovery and access, usage, and integration, and listed active and future mission support.

Andy noted that EOSDIS is phasing out FTP. Kristi offered to present at WGISS-40 what is happening with FTP at USGS.

**Action WGISS-39-08**: Andy Mitchell to consider a session for WGISS-40 on the retirement of FTP.

## Centre National d’Etudes Spatiales (CNES)

Jerome Gasperi gave a presentation of the French Exploitation Platform for the Sentinels data (PEPS). In March ESA and CNES signed a collaborative agreement on managing and accessing Sentinel data. PEPS is France’s official access point to the Sentinels data. Eight Sentinel products are retrieved from the ESA hub per minute and stored at the CNES data center. Product metadata are semantically enhanced by the iTag automatic process, based on the REST web services. Jerome gave a demonstration.

Satoko asked how the land cover percentage is calculated. Jerome replied that footprints are analyzed during the ingestion of the metadata, and various factors allow characterization of the product, using the EC Global Land Cover from 2010.

## European Space Agency (ESA)

Mirko Albani gave a presentation of the Copernicus space component data access. He began listing the Copernicus dedicated missions and their operational milestones: Sentinel-1 (SAR imaging), Sentinel-2 (multi-spectral imaging), Sentinel-3 (ocean and global land monitoring), Sentinel-4 (geostationary atmospheric), Sentinel-5 (low-orbit atmospheric), and Sentinel-6 (low inclination altimetry).

Mirko explained details of the Copernicus data policy, acknowledging that Copernicus data access is a major undertaking, considering the requirement for systematic processing and availability of all the acquired data, corresponding to a sustained generation rate (24h/7d) of a continuous stream of 500 Mb/s of user products (Sentinel-1,-2,-3 A series).

The Copernicus Space Component Ground Segment data access is ensuring that all Sentinel products are accessible to all users online. Access to Sentinel products is made available via dedicated data hubs. In addition, access to full Sentinel long-term archive is being made available to all users online ([www.sentinels.copernicus.eu](http://www.sentinels.copernicus.eu)). There are four main Sentinel data access hubs: Scientific and other access, access for Copernicus services, access for international agreements, and access for collaborative ground segment. Richard commented that the download limitation is three weeks. Mirko listed the Copernicus core users.

The Copernicus Space Component Data Access (CSCDA) System provides since 2007 comprehensive and coordinated access to EO data products from multiple satellites to Copernicus space data users across Europe. Transition from Data Warehouse phase 1 to phase 2 began on 1 April 2015, for the provision of data from fleet of CCMs ensured over next 6 years. The CDS will continue to guarantee access to Contributing Missions and Sentinels to Copernicus Services with strict Quality of Service requirements.

The Data Access Hub status includes simple self-registration; routine data flow opened following IOCR in October 2015. A rolling archive of at least last two months of data, and all data produced is online are available, with APIs provided for automatic downloads via scripts. There is a quota restriction of two concurrent downloads to ensure bandwidth availability for all users.

The initial CollGS Data Centre infrastructure is in operations as of end January, with seven accounts enabled, one per MS with CollGS agreements formalized. ESA front-end software DHuS features Open Source Framework management; open source code will be available in the middle of 2015.

Mirko listed the main services to users, and showed trends in number of users and downloads (6700 registered users, 115000 products available for download). He noted that the total mission production has been downloaded more than 6 times. Europe and North America are hosting the most active user communities.

The opening of the Sentinel-1 data flow to all users took place on 3rd October. Data can be accessed from: <https://sentinel.esa.int>. The Sentinel-1A spacecraft is in a stable state, operating in Nominal Mission Mode (NMM), with all sub-systems working on prime units. The Flight Operations Segment (FOS), ensuring the monitoring, control and commanding of the satellite, is operating nominally. Orbit control maneuvers are performed typically once a week. X-Band data acquisitions are routinely performed over Matera, Svalbard and Maspalomas X-band core stations. The acquired data are circulated within the PDGS, systematically processed to Level-0 and Level-1 products and archived at PACs (UK and Germany). Level-2 product operational qualification is on-going. Sentinel-1A is contributing to the scientific analysis of the M7.8 Nepal earthquake that occurred on 25 April.

The CSC data access is based on three main pillars: Latest data availability from online rolling archives; access to long term data archive; reduce download needs – bring the users to the data. The CSC data access is in continuous evolution to adapt to evolving user scenario and needs, to introduce the latest IT technologies, and to implement a continuous performance improvement process. The CSC data access enhancements are coordinated with the Commission via a change management process defined in the EU-ESA Copernicus Agreement. They are required to guarantee agreed performances, measured via specific KPIs, and shall not compromise the overall CSCDA reliability, performance and network security.

Short term CSC data access enhancements include gradual extension of access to the Sentinel long term data archive to all users as of mid-2015, with enhanced connectivity to GEANT network via dedicated 10 Gbps line. Mid-term CSC data access enhancements include integration of Sentinels -2, -3, and Sentinel-5p, integration of hosted processing capabilities. Long term CSC data access enhancements include roll-out of research and support services, federated user management pilots.

Mirko said that anyone can access the archive online; all the data is in the archive.

## UK Space Agency (UKSA)

Wyn Cudlip presented a report on the Catapult Satellite Applications (new UK Technology Innovation Centre). He explained the structure of the Catapult Network, which is one of seven independent centres, part-funded by the Technology Strategy Board driving economic growth through commercialisation of research. He explained its relationship with other organizations. The goal is for stakeholders connecting SMEs with end users to produce applications that solve real problems and that have market potential. Engagement can be by contract, collaborative work, or self-funded projects. Catapult provides end-to-end infrastructure that enables prototyping and demonstrations, access to test-facilities and expertise, data integration, and facilitates access to existing satellite infrastructure, communication networks and data sources. Wyn listed recent key collaborations.

The first downstream markets they are focusing on are transport, security and civil protection, climate energy and natural resources and internet of things. It is important to connect satellite applications with downstream markets, and they are finding that some of the needs are very simple projects that are very useful and easy to achieve. Wyn showed a number of examples using Landsat scenes.

Jerome asked about the classification of terrain (since a great deal of in-situ data is required for verification). Wyn replied that they use the forest classifier technique.

Wyn reported that the UK Collaborative Ground Segment for Sentinel data is funded by the UK Space Agency. Public sector users and innovators with a national interface to the Copernicus data.

## Global Spatial Data Infrastructure (GSDI) Association

Gábor Remetey-Fülöpp, Secretary-general, HUNAGI, presented an update on the activities of the Global Spatial Data Infrastructure (GSDI) Association. He mentioned updates in the personnel of the GSDI management and high-level cooperation with learned societies of the geospatial world. Gabor listed GSDI Activities in Marine (Coastal and Sea) Spatial Data Infrastructure

GSDI was asked to run a Round Table Discussion on Open and Big Data at the Defense Geospatial Intelligence Conference in London. GSDI participation at the 4th GEO Workshop on “*Concepts, Technologies, Systems and Users of the Next GEOSS,*” organised by the GEOSS Science and Technology Stakeholder Network (STSN) in Norfolk VA (March, 2015) <http://www.gstss.org/2015_Norfolk_4th>. GSDI contributed to the EU BYTE Project “The Big data roadmap and cross-disciplinary community for addressing societal Externalities” [www.Byte-project.eu](http://www.Byte-project.eu) at a brainstorming workshop in Vienna on 13 April, 2015 participated by GEO experts and GEO Sec Osamu Ochiai.

Gabor listed references to newsletters and publication, and other dissemination and promotion activities. GSDI Regional level member activities (Example: update on EUROGI) related to the Earth Observation were listed, as well as GSDI National-level member activities. These include EURISY Regional event supported by HUNAGI and HUNSPACE, and an International Conference on GEOSS organized by Óbuda University, Institute of Geoinformation. Gabor announced that Hungary became 22nd full member of ESA. They continue Involvement in NASA World Wide Europa Challenge 2015, and in innovative development by individual members, promotion of EO-Related Applications Calls.

He concluded noting that interoperable spatial data infrastructures and related services are enabling tools for EO applications. GSDI and its regional and country level members are playing an active role not only serving and supporting EO applications, but also providing awareness raising and user feedback by arranging international networking, regional projects, conference sessions, thematic workshops and promoting challenges from local to regional in EO. The Danube Region Data Service Infrastructure project offers excellent opportunities to investigate the feasibility to launch an ESA-supported pilot, with the aim of the establishment of a Danube Data Cube based on WGISS experiences but with the feature of cross-border environment. A draft note on this vision will be submitted to Mirko Albani of ESA ESRIN after 20 May 2015.

# WGISS Projects

## CEOS Water Portal Project

Satoko Miura opened the Water Portal Project session. She noted that the plan is to hand over the operation of the water portal to a third party in April 2016, so the WGISS project will conclude at that time.

### Operation Status

Shinichi Sekioka gave a presentation on the status of the CEOS Water Portal. He began saying that the CEOS Water Portal is a distributed data system component of the DIAS (Data Integrated Analysis System) Program to provide an easy-access service to users to a whole variety of hydrological data and water relevant data and to connect data centers, scientists and users. Multiple types of data are available such as in-situ, satellite and model output data. Services include dataset search (category and map search) of 11 data centers and two catalog broker systems. Dataset access depends on the server side function, and includes data subset (time, variables) and data download/format conversion (NetCDF, ASCII, GRIB-Model output only). Use case registration/browsing are also available. Shinichi listed the data partners and available data sets.

A new architecture is being developed to simplify the task of adding data partners, to provide easier operation, and integrated operation flow (search -> download). Shinichi showed a diagram of the new system architecture

The new CEOS Water Portal site is available from May 12th, 2015; the number of registration and visitors are on the rise. Registered users are about 290. The search interface with GEO DAB for GEOWOW river discharge data is in place.

Activity plan for FY2015 includes integration with DIAS catalog system, development of a common dataset level catalog database, and using GI-cat software. Plans also include changing the user authentication function to SSO enablement and alignment with DIAS authentication service. DIAS will take over the Water Portal system operation at FY2016.

### Demonstration

Kaori Kuroiwa gave a demonstration of the new version of the water portal that is now in operation, focusing on search capability for GEO DAB.

### Discussion

Wyn asked if there is access to services and is any processing capability supported; Satoko replied that access is only to the data and there is a subsetting feature for some datasets. Michael asked if the portal allows search by GCMD keyword; Satoko confirmed that it does.

Andy commented that three NASA systems are on the partner list including ECHO and DAACs. There may be an overlap as ECHO (future CMR) has all the DAAC server data. NASA lists the DAACs as they are individually OPeNDAP servers which provide actual data, whereas ECHO is listed as a broader catalog service. Satoko suggested that they may need to look into the overlaps. Andy asked if there is any data they want that they do not have; Satoko replied that their user group has been requesting NCDC datasets. This data addition will be a focus this fiscal year.

Richard asked which SSO technology is used; Satoko replied that it is CAS. Richard commented that GI-Cat cannot handle more than 1 million granules, and wondered what they would do when they reach that target. Yoshiyuki replied that they successfully ingested less than a million entries, but failed to add more. The search did not work correctly with more data. Richard asked if they are using Hyrax or THREDDS servers; Yoshi replied that they collect catalog records from the OPeNDAP servers in order to allow for the dataset discovery on the portal.

Satoko said she plans to close the WGISS Water Portal Project at WGISS-41.

## Disasters Recovery Observatory (RO) Project

Richard Moreno introduced the Recovery Observatory session. He stated that WGISS still has time to make changes or propose functionality for the Recovery Observatory, which has been presented to the WGDisasters. The WGDisasters members are pleased with it, and made recommendations which have been implemented.

While satellite imagery is used on an ad hoc basis after many disasters to support damage assessment and track recovery efforts, there is currently no system to support the coordinated acquisition of data and to facilitate its access. The concept of a “Recovery Observatory” was initially born from seeing huge quantities of Earth observation data that are made freely available following major disasters to many different users. The goal is to increase satellite Earth Observation contribution to all the Disaster Management Cycle phases. The WGDisasters has not yet triggered an RO.

### Main Features

Mathieu Gond gave a presentation describing the progress of the RO Project. The main features are a collaborative web-based workspace for groups of users to provide groups and communities with collaborative tools, with product search capabilities (multi-criteria and map-based search), and full display of full resolution geo-referenced images. Earth Observation products such as high and very high resolution optical imagery, radar imagery (X, C, L-bands), and airborne data are available for download.

The RO also has community contribution capability for in situ data, ground truth, maps, added-value products, reports, and publications. The web portal features editorial content, news, partners, and thematic-based collaborative groups with security management. Products footprints and full resolution products are displayed on a map, as well as lists of available products, including details on metadata, quicklook and footprint. The tool has multi-criteria search using OpenSearch parameters, and performed by RESTo but rendered by Drupal.

### Technologies and COTS

Mathieu listed the main technologies (PHP, JavaSpring, HTML5, CSS3, JavaScript, and OpenLayer), and COTS used (Drupal 7 with Acquia Commons, PostgreSQL, PostGIS, MapServer, GDAL, RESTo, iTag, and Mapshup). The Open Source Content Framework system is written in PHP, and includes basic features (user account, menu, page customization, and system administration), and plugins to extend the core, and is fully customizable using themes

### Technical issues

Mathieu discussed technical issues encountered. The ingestion unit is a scalable system, an archive containing the metadata of the data and a set of product archives. The Java JMS queues based on architecture.

Drupal and RESTo require coupling, since RESTo stores the products and the metadata and defines the download rights, while Drupal stores the articles and documents. Communication of these two is achieved with user session sharing, product metadata duplication in Drupal, and multi-criteria search form with map coupling.

### Future Features

Future features planned for the RO are to expand user contribution, license management, and user permission search results filtering. Any authenticated user will be allowed to add a product contribution to the system; contributions (full ingestion and supported product types) may be moderated depending on the user’s trust certificate. A license management feature is also planned; each product will have its own license, and user will have to sign this license in order to download a product. Currently all products that map with a given search query are returned, but in the near future, depending on the user permissions, some results will be hidden.

### Live Demonstration

Mathieu gave a live demonstration of the RO. He began with an overview of the home page, which includes a posting site for blogging and a news tab. He demonstrated how to do a search among the holdings of the RO, showing the details of the product, including the metadata and terrain type. He also demonstrated search features with multi-criteria form such as bounding box, date, platform, cloud/terrain/ice cover, adding that a partners’ page is available. He noted that it is easy for someone to view the metadata by scrolling down. The idea behind the data acquisition plan page is to display a spreadsheet on the screen.

In response to a question from Wyn, Mathieu indicated that they do not have a way of dealing with multiple images, stacking them up. Chaoliang asked if all the footprints have already been tagged a return a name or place; Jerome will explain this more clearly in his presentation.

Richard commented that in the search box the user can put search criteria in sentence form, and the system will interpret it and return a result. Andy asked if they use the controlled list of keywords. Richard said they have their own list, but are interested to compare it to the IDN keywords.

Yonsook asked which data is being used; Richard replied that this will be decided when the disaster is defined; the objective is not to have only space data, but also photos, in situ data and documents that can be useful for everyone. The restrictions are that the data is geolocated and is time bounded. Andy asked if it would be useful to have a subscription service so that the data can be added quickly, and to feed the system automatically. Richard replied that currently the data is inserted by the community, but a subscription service would be good. Yonsook asked if they are working through any user groups; Richard replied that they are currently working through the KalHaiti user group, but it will not be the same group for flood/volcano/earthquake disasters; the user group will have to be adapted to the type of disaster. They are working very closely with the WGDisasters, and have not received any new use cases.

Richard asked for input from WGISS regarding the roadmap Mathieu has shown; Andy felt that it would be best to stay with the current plan, and to stay connected to WGDisasters. The original request was for a single RO, but the WGDisasters is saying they want to have more than one instance. It is not a huge system so it can be deployed multiple times.

Richard concluded saying that the KalHaiti group is currently reviewing the RO, and will replace their system with it, since it allows the community to comment and collaborate. It is a different approach; comments and information are given by the community. At WGISS-40 the team will show the final version and the conclusion of the project.

# Interest Groups

## Technology Exploration Interest Group

Andy Mitchell and Satoko Miura chaired the Technology Exploration Interest Group session.

### Cloud / Big Data Agency Update

Andy noted that WGISS has reported on cloud computing and big data issues at past meetings, and that ROSCOSMOS and NASA were prepared to present additional agency information on the topic.

#### The Basic Products Bank Service – ROSCOSMOS

Tamara Ganina introduced the ROSCOSMOS Basic Earth Observation Data Products Bank (BPB). It is the result of special processing of standard remote sensing data product, and a basis for multipurpose thematic interpretation for solving the tasks of interagency and regional customers. Today the BPB contains datasets from which the different characteristics of region, phenomena, and/or objects observed may be derived. They allow monitoring of long-term variability objects, constructing maps of geophysical parameters, comparing data from various sources of satellite observations, validation of remotely retrieved geophysical parameters describing physical objects, and acquisition of source and reliable data needed for models of various phenomenon and process development.

Tamara displayed a diagram of the hierarchy of BPB categories, primary basic products, secondary basic products and composite images, and she described each of these.

The objectives of the bank (BPB) are the use of raw data from the Russian in-orbit satellite constellation for addressing socio-economic problems. The agencies need to develop the technologies of basic data product generation autonomously at great expense. The ultimate benefit is to have a federal public BPB that provides the agency-level and regional systems with multipurpose source data to solve the thematic tasks of space monitoring without the need for developing and operating the relevant technologies themselves.

The set of technologies can be considered as a geo-information web service, where customers can place an order. The back-end part of service generates the secondary basic products and composite images. Generation of the high-level basic products is realized based on the primary products stored in the long-term archive. High-level basic products can be downloaded from the HTTP-links.

The central component of the BPP is a resource administration component that provides the interoperation between other components, such as Archive and Database component, BP generation, BP Verification and BP Distribution components. Moreover, BPB provides a geographically distributed data processing model and includes the central and regional segments. The regional segments include local long-term data archive, BP generation component, subcomponents of the resource administration component.

The resource administration component provides the Map/Reduce programming model and includes a dispatcher that performs the high-level distribution for segments, a controller that provides the distribution for applied problems, and a worker that performs the low-level console application initiation. Communication between modules is realized via TCP/IP protocol.

The object model, including objects classes and the procedures of various type semantic network model are developed. The set of classes and communication types of the object model is based on «Common geodatabase metadata profile». The database was deployed with use of Oracle high-availability cluster technology that provides scalability and BPB database operational continuity. Unified basic products include raster data (presented in cartographic projection, tiff presentation format, 16-bit data presentation) and metadata (satellite, camera, image, and raster data description), packed into zip format.

Tamara showed an example of a composite image, generated from Meteor-M2 KMSS sensor and Terra MODIS sensor. The metadata contains the description of initial data from which composite is generated and attached to the image. Tamara also described the basic product verification component, and the distribution component, concluding that BPB has two mechanisms of basic data bank access. The access is Personal (with the use of interface tools for catalog products searching and making an order, the order state monitoring and downloading of the generated products) and Subscription (in accordance with the regulations specifying the requirements for products and means of access to them by main federal consumers).

Full access to BPB technologies will be possible from 2016 onwards.

Chaoliang and Wyn asked if the BPB includes magnetic and Sentinel data; it does not yet but the hope is to do so in the future. Andy asked if the BPB verification keeps a copy; Tamara replied that they use the standards of verification, and are capturing metrics on the users. For the individual users they need to understand what product the users need. Richard commented that they use an Oracle database cluster, and wondered why they chose Oracle; Tamara replied that it is because of the high availability.

#### NASA

Andy Mitchell gave a presentation on the NASA-EOSDIS Next Generation Application Platform (NGAP). Each application deployment (re)creates this process and they all deal with the same core problems. Current application deployment can be time-consuming and expensive to scale and to maintain. The question is how to provide cross cutting resource management. NASA set out to perform a Trade Study that investigates a scalable, flexible application platform solution that offers the cost benefits of hardware consolidation with the safety and security of application sandboxing and resource management.

Andy observed that procurement of hardware can take time, setting up VM’s requires a team of system administrators, developers and database administrators; infrastructure is heavily reliant on third-party approvals. Initial network setup and network changes to make an application available for access can take a week to occur, and analysis of monitoring logs and patching takes several hours per week, every week. Failover and backups are only implemented for applications with a high need.

The solution is to select the best ideas, and use shared resources. After investigating COTS solutions and selecting to the best fit that meets NASA’s fault tolerance, monitoring, metric, and security demands, the result of this investigation was to provide a cloud-based Platform-as-a-Service for ESDIS applications: The next generation application platform (NGAP).

NASA is currently developing a prototype small scale implementation (private and public cloud) of a highly available, scalable platform that can host future NASA Earth science data system services. The beta release of the NGAP will run in NASA’s Computing Services Service Office (CSSO) General Purpose Managed Cloud Environment (GP-MCE). It will also be integrated with NASA’s Earthdata Code Collaborative (ECC) to allow for builds and releases.

NASA Sentinel Mirror is leveraging existing EOSDIS capabilities to serve US scientists and NASA is developing a data mirror archive for Sentinel 1, 3, 5P in the U.S.A. The single network interface relieves bandwidth load on European networks; long term archival and end user distribution by DAACs (e.g. Sentinal-1 by ASF DAAC). It provides metric reports back to the EC/ESA on product distribution and usage, and leverages the entire suite of NASA’s EOSDIS capabilities including capturing and reporting metrics on distribution and usage of Sentinel products by U.S. scientists.

The NASA Sentinel Gateway Target ensures the reliable archive of the golden copy of Sentinel data in Washington, the distribution to the appropriate DAAC(s), and makes that data available in AWS S3 storage. The approach will not be limited to AWS but that is assumed to be the initial target to support NGAP.

Satoko noted that JAXA is also investigating system usage, and having their own cloud is also a possibility. Satoko suggested that software upgrades and software porting is one of the biggest challenges and that with commercial cloud, the OS can be upgraded without notification to the users. JAXA has a supercomputer system, but is now doing an evaluation; NASA also has a supercomputer system that is being used to re-process their data. Andy said that NASA’s office does not know what the cloud will look like in a few years.

One advantage with using the cloud for distribution is that only the most popular datasets would be stored in the cloud. Richard asked if the user pays for the cloud processing; Satoko replied that it is one of the things they are studying, and have a use case for it. Currently they are distributing Sentinel from the spinning disk.

Satoko asked if they have any publicly available report of the study conducted by NASA.

Mirko noted that ESA is using public cloud for disseminating data, and that the applications are stored on the cloud. ESA is also following an approach to reprocess as much as possible and putting the data online, and they are also purchasing computing time on the cloud, which simplifies internal infrastructure.

**Action WGISS-39-09**: Mirko to prepare a presentation for WGISS-40 on ESA’s cloud processing.

### Data Management Plan

Andy Mitchell introduced a session on data management.

#### GEO Data Management Principles Task Force

Andy Mitchell gave highlights of Data Management from a GEO perspective. The GEO Infrastructure Implementation Board (IIB) view of data management is that it is progressing on a number of fronts (e.g. processing, validation, quality control, modelling, and visualization); however the development and implementation of best practices for data management varies a lot from country to country and organization. Because of this, GEO established a Task Force to draft GEOSS Data Management Principles for submission to the GEO-XI Plenary (November 2014). The data management principles cover the entire data life cycle from planning, to acquisition, quality assurance, documentation, access, archiving, preservation and answering user needs.

The DMP TF convenes to work with GEO collectively, including the Implementation Boards, the DSWG and related WP Tasks to develop principles for:

* ensuring that data are properly managed, accessible, archived and long term preserved,
* ensuring that data are properly documented, quality controlled and quality assessed, delivered, and updated in ways that facilitate access and re-use of information made available through the GCI,
* facilitating the link between user needs and data availability, especially with regard to the needs of users from developing countries (e.g. by identifying existing sources of requirements already approved by the relevant user community), and
* facilitating interoperability of GEOSS data resources by promoting a progressive harmonization/standardization of content (data models, thesauri, coding list, ..) and dissemination and usage rights in order to facilitate their re-use at global or regional scales.

The proposed data management principles are:

* **Discoverability:** Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions, including licenses, will be clearly indicated.
* **Accessibility:** Data will be openly accessible with minimum delay and cost. Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.
* **Usability:** Data should be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards. Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international standards. Data will include provenance metadata indicating the origin and processing history of raw observations and derived products, to ensure full traceability of the product chain. Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.
* **Preservation:** Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures. Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.
* **Curation:** Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate; where applicable this shall follow established and agreed procedures. Data will be assigned appropriate persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

Although the TF has been set up to work until the GEO-XI Plenary and does not cover implementation guidelines, some “principles” already emerged when discussing the comments receive. These could be the initial basis to further develop preliminary guidelines for implementation:

* Data management principles are not legally binding.
* When data management practices already exist, they should be respected.
* Adoption and compliance with the DM principles may be gradual. Data providers should have the opportunity to decide at which level they can afford the costs to be fully compliant.

#### Earth Science Information Partners (ESIP) Guidelines

Andy Mitchell stated that the Federation of Earth Science Information Partners (or ESIP Federation) developed an “Elements of a Data Management Plan” module for their Data Management for Scientists Short Course. The module was authored by Ruth Duerr from the National Snow and Ice Data Center at the University of Colorado in Boulder.

Elements of a data management plan are:

* “First, your data management plan needs to describe the materials you will be producing during the course of your project. What kinds of data will there be? How much of each kind? Who might be interested in using your data and what other information will they need in order to use them properly? In fact, what other information are you going to need to be able to re-use these data a few years from now?”
* “Second, once you know what you are going to produce, you need to decide how you are going to organize them. What filenames, directory structures, metadata and data formats will you use? How will all of this be documented?”
* “Third, what data do you plan to share? How will people access them? When will they have access -- immediately or after some time period? Can they redistribute the data? Can they create derived products? Are there any ethical, legal, or other restrictions on access and use, perhaps to protect personal privacy, endangered species or sensitive sites? If so, how will you handle those restrictions?”
* “Fourth, how are you going to ensure that your data are safe and haven’t been tampered with, both during your project and after it? Will the data go to a long-term repository? If so, what arrangements have you made with the repository? You have planned to make arrangements, haven’t you? They do know you are expecting them to take your data and that you have planned to meet all of their requirements for data deposit, right? Speaking as an employee of a data center, the last thing most repositories want is to have someone knock on the door at the end of a project expecting to just drop off the data. That situation rarely turns out well.”
* “Last and perhaps most importantly, whose job is it to make sure that all of this is actually carried out? For example, who will create the metadata and other documentation? Who will make sure that any sensitive data are secure and that only authorized personnel have access to them? Who will ensure that your data are backed up, that they haven’t been corrupted, and that multiple copies are available from distributed locations?”

#### USA Open Data Policy

Andy Mitchell stated that in 2013, the US government released a memorandum ‘Making Open and Machine Readable the New Default for Government Information’. This established a framework to help institutionalize the principles of effective information management at each stage of the information's life cycle to promote interoperability and openness. The US President also released a memorandum on ‘Transparency and Open Government’ instructing US agencies to take specific actions to implement the principles of transparency, participation, and collaboration. The Office of Management and Budget's (OMB) Open Government Directive required agencies to expand access to information by making it available online in open formats. In addition, the US Government launched [https://Data.gov](https://data.gov/), an online platform designed to increase access to Federal datasets.

#### USGS Data Management Plan

Andy Mitchell presented information on the US Geological Survey (USGS) data management plan. Earlier this year, the USGS released an Instructional Memorandum (IM) that provides interim policy for establishing a USGS data management foundation following a data lifecycle. An overview of the results to be achieved by the science data lifecycle elements are as follows:

* Plan: The overall project work plan of every research project must include planning for data management.
* Acquire: Methods and techniques for acquiring research data are planned and documented to ensure that scientific findings are verifiable.
* Process: Data processing denotes those actions or steps performed to verify, organize, transform, repair, integrate, and produce data in an appropriate output form for subsequent use.
* Analyze: Analysis involves actions and methods performed that help describe facts, detect patterns, develop explanations, and test hypotheses (for example, statistical data analysis, computational modeling, and interpretation of results).
* Preserve: Preservation includes actions and procedures that are performed to ensure that data are retained and accessible consistent with the USGS Records Disposition Schedules and other applicable regulations.
* Publish/Share: USGS scientific data may be released or disseminated in a variety of ways, for example in datasets and databases, software, and other information products including USGS series publications outside publications and USGS Web pages.
* Describe (Metadata, Documentation): Metadata describe USGS scientific data and how they were collected and processed, and are essential for reproducibility of research results.
* Manage Quality: Data management activities (including use of standard methods and best practice techniques) are performed in a consistent, objective, and replicable manner to help ensure that high-quality and verifiable results are achieved.
* Backup and Secure: During all data management processes, backup copies of data are created to protect against loss that can result from events such as human error, hardware failure, computer viruses, power failure, or natural disaster.

#### NASA Data Management Plan

Andy Mitchell presented information on the NASA data management plan, whose purpose is to ensure uniform and complete implementation of the principles embodied in the NASA Earth science data policy. The following programmatic data system requirements have been created to be utilized in all satellite mission Level 1 Requirements. Andy detailed the science data management requirements, and the science data requirements.

“All terms and conditions of the transfer of data products and associated information to the NASA Earth Science Division (ESD)-­assigned data center shall be documented in a Data Management Plan that has been approved by the Earth Science Data and Information System Project.” The purpose of the Data Management Plan (DMP) is to address the management of data from Earth science Missions, from the time of their data collection/observation, to their entry into permanent archives.

Activities to be included in the plan include determination of list of products, data and metadata formats, development of ICDs and OAs with Science Data Centers, interface testing, end-­ to-­end data flow testing, planned start and end of data collection, start of flow of data to Science Data Centers, start of public access to data, planned start and end of reprocessing, and assembly of archival information package.

#### Data Management Plan Discussion

Richard noted that WGISS is PoC of GEO IN-02-C1. WGISS can offer to provide a number of documents. Mirko added that WGISS members are also co-chairs of the DMTF. It is expected that they will assign teams of 2-3 people for each data management principle. The guide will not be technical. Instead it will point to the appropriate technical documents.

### Open Source Software (OSS)

Andy Mitchell presented background information on Open Source Software, which is any computer software distributed under a license which allows users to change and share the software freely. Open source software is required to have its source code freely available and end-users have the right to modify and redistribute the software to others.

WGISS benefits from OSS since OSS is much better at adhering to open standards than proprietary software is. Because interoperability with other agencies is valued, open source software is definitely the best approach. This allows the agencies not to be limited by proprietary data formats.

Barriers to use of OSS are that many of the traditional software licensing mechanisms are viewed as a potential barrier to software re-use. Traditional licensing typically requires the re-user to negotiate usage terms and conditions with the intellectual property owner every time that they want to re-use something. It is worth noting that open source licensing is not appropriate for ALL software. For example, the presence of proprietary code or export control restrictions may be valid reasons for choosing not to go open source.

Discussion points for WGISS:

* How do you use open-source software in daily practice?
* What repository do you use? What license?
* How does management at your agency encourage open source business practices?
* What is your flagship open sourced application? What problems did you encounter along the way? For NASA, GIBS is probably the most popular one; they are trying to get ECHO and CMR into open source.
* Does your agency have intellectual property issues involving copyrights or code provenance? Ignoring legal issues with software is costly.
* Does your agency have legal controls and processes in place to deal with open source software?
* How easy or hard has it been to integrate open source software in your agency?
* Does the license suit all your future plans for the code? The GNU General Public License (GNU GPL or GPL) is the most widely used free software license, which guarantees end users (individuals, organizations, companies) the freedoms to use, study, share (copy), and modify the software. Some open source licenses allow for free use in commercial applications and others do not. Some open source licenses specify some restrictions when you host software-as-a-service.
* Is your code (or the code you are using) well architected and implemented? Experts should asses the quality of the code and quality of the documentation and user interface.

Andy said that in the U.S. they face great difficulties to share source code even within national agencies; everything has to go to the Patent Office. Technically, all code is open and available, but it is bureaucracy that provides the hurdles.

Mirko said that at ESA they try to develop open source software and toolboxes, and are also trying to move to open source licenses. It is easier if it is open source from inception. Mirko added that on Sentinel toolboxes they are using GitHub, and they may have many other open source.

Wyn asked if there is a distinction between giving for use, and giving for further development. It depends on the lawyers’ interpretations. Support for open source can have a huge price tag. JAXA is in a similar situation – they can give the executable, but not the source code, which may have been developed by many different vendors.

Richard said when CNES develops, they need to valorize the code so it can be sold to the contractors. CNES wants science laboratories to be able to supplement the code, and be reused by other laboratories. The objective is to be more open, while keeping control of the software.

If any member knows of applicable OSS, WGISS would be interested to have the information on the WGISS OSS web page. It was agreed that the Open Source page should be organized according to the five GEO Data Management Principles.

**Action WGISS-39-10**: Technology Exploration Interest Group to plan a session on Open Source Software practices inviting each agency to present at WGISS-40.

### OpenDAP

David Fulker\* gave a presentation on recent advances in remote data access via OpenDAP web services. He began explaining the concepts of OpenDAP, a data access protocol. Its internal data model has few data types for the sake of simplicity. Types are domain-neutral but flexible. These types support many domain-specific needs. OpenDAP services function as middleware. Data ingest via encoding-specific adapters, and multiple response encodings. He presented an architectural overview of Hyrax, a widely-used DAP server.

DAP-based subset selection (from arrays, tables), selects variables by name, rows of a table via column-specific value constraints, and sub-arrays by constraining their indices. Index-based subsetting is excellent if the desired subset is a bounding box parallel to source array. But it is less useful when subset selection is not based on domain coordinates, or source is not organized as coordinate-mapped arrays, or desired subset is polygonal or is skewed relative to source-array orientation.

Next David discussed recent enhancements in OpenDAP services. The main NASA motivations for OpenDAP enhancements are easier software builds and better documentation, authentication of data users, more response encodings, and requesting DAP ops on many granules at once.

Hyrax-install complexity was once a barrier to use. Key accomplishments toward simplification are learning that adding modules does not increase the package count. Source build is now just three instead of 18 packages.

David listed progress on enhancing the OpenDAP website and documentation and on user authentication, and progress on multi-granule aggregation, and explained DAP output-encoding extensions. He described the OGC protocol for WMS web mapping service, and gave a demonstration of the interoperability at a very sophisticated level. This can be done on many file types, and only requires specific encoding handlers, leveraging the possibility for a large variety of clients.

David noted that they demonstrated NASA (HDF5) files🡪 OpenDAP🡪WMS🡪 Google Earth, and commented that it seems unlikely that either Google Earth engineers anticipated reading HDF5, or NASA engineers planned to display data on Google Earth. This suggests a definition for interoperability as “supporting unanticipated uses”.

Satoko asked if this is operational. David said it is, and they hope to have all their data centers on OpenDAP. She also asked how many scenes (data volume) can be handled by the OpenDAP interface, how scalable it is. David said that their experience is that it scales very well, and it is scalable beyond what it is typically used, supporting many concurrent users.

Kristi asked about URL nomenclature. David replied that in principle the intention is for it to be consistent with the PI approach. It allows for conveying identifiers for subsets. It is their desire to have it consistent, but do not know if they are adhering to any protocols that have been established. Andy said that the URL is based on how the server is set up. So if changes in the server occur, then there will be changes. That is a policy rather than a technical question.

Yoshi asked if there are two choices, one using Hyrax, and the other THREDDS. David replied that Unidata is the provider of the THREDDS, and they have worked very hard to keep them consistent. They are quite sure that authentication service works with THREDDS, you can choose either one.

Richard asked if it is suitable for optical data, radar data, and data in TIFF format. David said yes, they support TIFF and have adapters so they can support GEOTIFF as input as well.

### Data Cube Workshop

Two data cubes were presented: The GA data cube, and the CEOS Data Cube.

#### GA Data Cube

Simon Oliver and Jonathon Ross presented background on Geoscience Australia. GA has a historically strong relationship with the US Government on Landsat mission support, the NASA Systems Engineering Office for KenyaCube in support of GFOI/GEOGLAM, and is developing a Memorandum of Understanding with the European Commission/ESA, seeking to engage more closely with ESA regarding Copernicus.

As well as dealing with the explosion in data volume, variety and velocity, there is also an increasing expectation amongst users to have on-demand access to information products derived from data received in near-real-time. Data collection is dynamic: growing in time, and also subject to modification (existing data) and insertion (new data). The challenge is to enable:

* Attribution of exact observation time for key applications e.g. tides for shallow-water bathymetry, bare earth.
* Analysis of each observation in the time-series.
* Reliable comparison of observations over long periods of time, e.g. change detection, pattern analysis.
* Iteration and refinement of processes at continental scale.
* Rapid generation of results.

In response to these challenges, GA, along with collaborators, undertook a project to liberate the tape-based Landsat archive with the overall aim of reducing barriers to exploitation of the data resource. Key to this is providing a set of comparable measurements on an infrastructure that would allow rapid and repeatable analysis. The challenge of big spatial data has led to the development of the Australian Geoscience Data Cube. The AGDC is a common analytical framework composed of a series of data structures and tools which facilitate the organisation and analysis of large gridded data collections.

Australia Government recognised the need to invest in high performance computing research infrastructure. As a result GA and its key partners in collaborative science, CSIRO and the Bureau of Meteorology, have access to the country’s largest supercomputer. Without access to this sort of computing power and storage the work with the AGDC would not be possible. The National Computational Infrastructure (NCI) is an important part of the government’s eResearch infrastructure and brings together important. GA pays a relatively small contribution to be part of this consortium. The greatest benefit is that much of the important EO data is collocated and accessible on the same machine. The diversity of data present on the infrastructure provides for some exciting potential data fusion opportunities and it is not only Earth sciences as you can see.

Another key element of the solution is appropriately structuring data for analysis. Remote sensing data is typically both spatially and temporally irregular: i.e. image footprints vary between repeat observations, and successful acquisitions do not occur on a regular basis. 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 and Stack” method used to subdivide the data into spatially-regular, time-stamped, band-aggregated tiles which can be traversed as a dense temporal stack. Additionally, a solid scientific approach to data processing is essential, taking the data to comparable quality assured measurements. Simon mentioned some examples of the magnitude and complexity of the data.

The AGDC database provides indexing and filtering capability to enable attribute-based tile selection, and the AGDC API facilitates algorithm construction. It is written in Python and based on the open source Geospatial Data Abstraction Library/GDAL esp. Virtual Raster Transforms. The data grid specification based on the ANZLIC National Nested Grid Specification Guide – OGC DGGS SWG.

GA is heavily involved in the development of an OGC interoperability standard for discrete global grids through the establishment of a standards working group. The ultimate goal is to enable interoperability within grids in a grid system and between different implementations of DGGS.

The DGGS SWG will develop a version 1.0 implementation standard that includes a concise definition of DGGS as a spatial reference system. It also will include the essential properties of a conformant DGGS, the variability within these properties that classify types of DGGS, and elements of a spatial reference system identifier suitable for registering specific implementations of a DGGS.

Simplifying AGDC production tasks, Luigi enables construction of complex pipelines of long-running batch jobs by handling dependency resolution, workflow management, visualization etc. Conceptually, Luigi is similar to [GNUMake](http://www.gnu.org/software/make/) where certain tasks exist which may have dependencies on other tasks. Luigi takes care of a lot of the workflow management. GA have adapted Luigi to use the Message Passing Interface (MPI) for parallel processes execution on the HPC. Use of Luigi enables execution of embarrassingly parallel tasks associated with processing continent-wide processes across the 800+ AGDC tiles.

Simon reported that the code is now open-sourced on GitHub (<https://github.com/GeoscienceAustralia/agdc>), with a new release planned for May (<https://github.com/GeoscienceAustralia/agdc/releases>). They are ingesting new data collections using generic ingestion framework (e.g. MODIS), hardening remaining prototype code and optimising prototype DB schema. The team is also developing new APIs to support specific use case patterns, generic workflow tools to manage parallel processing (Luigi), and delivering basic WMS, WCS, WPS and WCPS web services while providing simple tools for cross-sensor interoperability (e.g. spectral matching/adjustment).

Being able to break AGDC analysis workflows down into a set of common modules and filters is an important part of simplifying access to processing in the API.

The Australian Geoscience Data Cube presents as a Common Analytical Platform which is beginning to realise the potential of Data-as-a-Service and is representative of a move away from distributed data holdings. Approaches based on moving the raw data to the user are not allowing the delivery of maximum value from the partner data. By consolidating data, the barriers to data integration seem to be fewer given that interoperability standards have typically been focussed towards enabling a variety of implementations. Common infrastructure offers unprecedented scope of high volume data fusion processes. Consolidated holdings offer unparalleled opportunities for data integration – by having data in one place and enabling users to interface to the data, all users will be able to realise the potential of having all of this data available in a consistent environment.

Yonsook asked what the challenges are for moving an existing data cube to an operational data cube. These will be learned with experience, and the key will be to abstract the system as much as possible for potential uses.

Gabor mentioned an Australian national program (Spatially Enabling Australia) and wondered if the data cube has any connection with this project. Simon replied that project is more to do with topographic mapping activities.

Andy asked how WGISS can support the AGDC, as there is great interest. The work has relied on a team of engineers, and additional resources are always welcome. The work has been focusing on the back end, but there is a lot of potential for work in web services and individual access. The source code is open source. Andy suggested a use case where WGISS can participate.

#### CEOS Data Cube Concept and Prototype

Brian Killough noted that there is a global need for Space Data Services; CEOS is leading several projects to demonstrate data services tools and build an architecture that supports a growing volume of data. It is expected that these data services pilot projects will provide a foundation for future operational systems that will be funded and managed by UN organizations and individual countries. CEOS agencies have a keen interest in this activity since it promotes the use and benefit of satellite data.

The CEOS vision is to provide the architecture and tools for countries to access and utilize satellite data and the training and capacity building necessary to utilize those tools and manage satellite data. CEOS will fund prototype projects to demonstrate tools and services for data access and utilization, and will consider low bandwidth constraints and the growing volume of satellite data in its architecture and tools, and CEOS will utilize advanced technologies in its architecture and tools, such as Data Services Platforms and Data Cubes based on cloud computing.

An example space data problem in Kenya is being considered, since Kenya could have 5TB of annual Landsat and Sentinel-2 data by 2017. Processing a scene takes about one minute in U.S. and one hour in Kenya. Downloading a scene has similar timings.

The data cube is a proven concept by Geoscience Australia and the Australian Space Agency (CSIRO); a multi-dimensional (space, time, data layers) data cube is an efficient and effective solution. A paradigm shift is scenes vs. pixels, products vs. unprocessed data. The Data Cube approach supports an infinite number of applications, makes it easier for users to access and use space-based data, and allows efficient time series analyses and data assimilation. It also puts the burden of the processing on the agencies. Not a solution for scientist, but a good solution for developing nations in need of products.

In the data cube approach, 2D (spatial) data is arranged temporally and spatially to allow flexible and efficient large-scale analysis. The “Dice and Stack” method is used to subdivide the data into spatially-regular (nested grid), time-stamped, band-aggregated layers which can be managed as dense temporal stacks. A standard Universal Transverse Mercator (UTM) Nested Grid is an effective method for subdividing scene-based data. The Nested Grid approach improves interoperability and sharing of data across domains and observation types with spatial consistency. Datasets from various missions can be subdivided with a grid spacing similar to their base pixel resolution.

The design is a flexible architecture that supports infinite user applications, increasing and diverse datasets, local or cloud-based deployment, and automated ingestion of new datasets. It supports Open Source (Apache v2.0) software to allow free and open access, Advanced Programming Interface (API) access, future data and capability growth, and commercial opportunities. Brian displayed a detailed Data Cube Architecture diagram and listed high-level data cube requirements:

1. Free and open access to software and APIs
2. Documented processes for Data Cube generation, new data ingestion, and application interfaces
3. Automated ingestion of new data layers and new data acquisitions
4. Support cloud-based or local deployment
5. Utilize “Analysis Ready” data products for satellite data layers
6. Enable user development of applications through flexible APIs
7. Utilize a UTM “nested grid” for multiple dataset interoperability and spatial consistency
8. Develop a baseline user interface that supports Data Cube statistics/analysis and optical image preparation (e.g., mosaics)
9. Architecture flexibility and standards to support multi-country Data Cubes

The User Interface Concept is a web based tool that utilizes Map Panel and a Tool Panel similar to many existing GIS tools. Brian listed the detailed requirements for the data cube user interface (image preparation, forest extent mapping, and change detection).

Brian gave as an example the Kenya Data Cube Project, led by NASA-SEO and the Australian Government (Geoscience Australia, CSIRO and the Department of the Environment). The project will involve a large number of stakeholders and funders (Australian Government, NASA, USGS, United Nations REDD+ and FAO, Gates Foundation, Clinton Foundation, SilvaCarbon). The project will involve a large number of CEOS groups: Space Agencies (satellites), NASA-SEO (data tools), and SDCG for GFOI, LSI-VC, WGISS (data archives), WGCapD (training) and GEOGLAM. The team met on March 9-11 to develop a vision, architecture and task plan for the project.

The first phase of the project (2015) will develop an initial Kenya Data Cube with Landsat and SRTM data layers, and demonstrate a user interface with statistics/analysis tools and image preparation capability (e.g. mosaics). This phase will also demonstrate an application including a land classification algorithm and a temporal change detection algorithm, and present results at the United Nations COP-21 Climate Change Meeting.

In the second phase (2016) the datasets (e.g., MODIS, SPOT, ALOS, Sentinel-1A, Sentinel-2A), regions (e.g., Colombia, East Africa), and applications (e.g., GFOI, SLEEK, GEOGLAM) will be expanded. The will also demonstrate automated data ingestion of new products and expand analysis tools, application platform and APIs.

In the third phase (2017) more data layers, regions, and applications will be added, as will continued optimization of ingestion, data cube processing, and APIs for expanded applications support, and demonstration of mobile application.

Yonsook asked for elaboration on the data cube contents. Brian replied that each dataset is layers within the cube; if the cube has spatial and time dimension the user can separate datasets but it is all one data cube (limited high resolution data, some good high resolution data, maybe some radar when cloud is a problem). Virtually it all sounds like one, but in actuality the different users will use different combinations of the building blocks. If someone has a precise application, and their application needs a specific set of data, they could build their own mini-cube with precisely what they need.

The challenge is to scale it up in terms of processing speed and storage. The end result may be several virtual mini cubes. Yonsook commented that the combination of data needed is going to be application-driven. Brian added that the goal is to teach people how to make their own data cubes (with proper documentation); if some of the most common ones are already built, they can follow the process, and utilize what is available. In the case of Kenya they specified exactly what they wanted, and that is what is being built.

#### WGISS Engagement with Data Cube

Brian Killough presented preliminary ideas on how WGISS can engage with the Data Cube:

1. Conduct a data format study to analyze the advantages and disadvantages of using GEOTIFF, NetCDF, and HDF formats in a Data Cube. Are there computational and data assimilation advantages? The SEO/GA can provide WGISS with a small data cube for testing.
2. Develop an automated data ingestion scheme that utilizes CWIC/FedEO to ingest new data products into the Data Cube infrastructure as functional layers. The SEO is working with USGS to create automated scripting and requests for Landsat data. Can this work for the Sentinel missions? No other missions have open access or would be desired as “regular” data layers.
3. Conduct a data compression study to compare different compression approaches (e.g., JPEG-2000, HDF-5) for the Data Cube storage. Consider initial Data Cube formation, delivery to countries and deployment, Data Cube re-gridding, and regular operations of algorithms.

Richard commented that the provision of a small data cube for testing by the SEO would be very interesting. Several agencies would be interested in performing studies on this. Yonsook said that an aid to understanding from a WGISS perspective would be trying to describe how the data could be used in a variety of applications, what kind of tools are needed. Some very high level use cases would really help understand and would really aid in bounding what the system is. Brian noted that a place where it is really useful is time series analyses.

Kristi commented that the data cube concept would require a change of the current data management paradigm. It is important to have a method to understand the data cubes to handle changes; adding data types to one mission results in significant complications. Brian said that the evolution of the data with time is a good point. Kristi noted that one of the biggest issues is how to combine the older data with the newer data, and added that applying standard algorithms and processing is critical.

Andy asked if there is a user community where WGISS could participate, and Brian pointed him to Simon and Jonathon, who seem to know who is doing data cubes. The concept in general is not new; a great service will be to put a little more structure on it, giving a common architecture and structure.

It was agreed that clear use cases are needed for WGISS to proceed.

**Action WGISS-39-11**: WGISS Exec to respond to SEO’s request to conduct a data format study to analyze the advantages and disadvantages of using GEOTIFF, NetCDF, and HDF formats in a Data Cube. Are there computational and data assimilation advantages? The SEO/GA can provide WGISS with a small data cube for testing.

**Action WGISS-39-12**: WGISS Exec to respond to SEO’s request to develop an automated data ingestion scheme that utilizes CWIC/FedEO to ingest new data products into the Data Cube infrastructure as functional layers. The SEO is working with USGS to create automated scripting and requests for Landsat data. Can this work for the Sentinel missions? No other missions have open access or would be desired as “regular” data layers.

**Action WGISS-39-13**: WGISS Exec to respond to SEO’s request to conduct a data compression study to compare different compression approaches (e.g., JPEG-2000, HDF-5) for the Data Cube storage. Consider initial Data Cube formation, delivery to countries and deployment, Data Cube re-gridding, and regular operations of algorithms.

## Data Stewardship Interest Group

### Overview and Updates

Mirko Albani introduced the Data Stewardship Interest Group (DSIG), which has been busy drafting common cross-agency best practices or guidelines on data management and stewardship for possible adoption by WGISS. A preliminary list of these documents is:

* Preservation Workflow
* Data Preservation Guidelines
* Preserved Data Set Content
* EO Data Generic Consolidation Process
* Purge Alert Procedure
* Persistent Identifiers
* Appraisal Procedure
* Archived data media transcription
* EO Data Stewardship Definitions

The group is also contributing 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. The group will also propose WGISS best-practices/guidelines drafted by DSIG for adoption in GEO and possibly as input to standardization bodies (e.g. OGC/CCSDS).

Mirko gave a progress report on GEOSS Data Management Principles Task Force (DMP-TF), established by the GEO Secretariat in May 2014. The theme of the TF is that each Earth observation is an irreplaceable asset to understand the past, describe the present, and forecast the future of the global integrated Earth system. The objective is to facilitate data management approaches that encompass a broad perspective of the observational data life cycle, from input through processing, archiving, and dissemination, including reprocessing, analysis and visualization of large volumes and diverse types of data.

Mirko listed the DMP-TF members. He also listed the data management principles already mentioned. He especially highlighted the Usability principles, and noted that the DMPs are included in the GEO Strategic Plan 2016-25. The detailed usability principles are:

DMP-3: Data will be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.

DMP-4: Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably formal structured metadata based on international or community-approved standards. To the extent possible, data will also be described in peer-reviewed publications referenced in the metadata record.

DMP-5: Data will include provenance metadata indicating the origin and processing history of raw observations and derived products, to ensure full traceability of the product chain.

DMP-6: Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

Next steps for the DMP-TF is to draft an implementation guide. The target audience are organizations that contribute data and services to GEOSS, and should be pitched at a strategic or managerial level. Expected contents, drafting approach, and timescale were also listed.

For WGISS, key principles are Discoverability, Accessibility, and Preservation:

Discoverability

DMP-1: Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions, including licenses, will be clearly indicated.

Accessibility

DMP-2: Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.

Preservation

DMP-7: Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.

DMP-8: Data and associated metadata held in data management systems will be periodically verified to ensure

Mirko also highlighted projects and conferences in which the DSIG participated, including the interagency project in Europe for AVHRR coherent data series generation (harmonized L0 data consolidation, processing, discovery and access), the **SCIence Data Infrastructure for Preservation with focus on Earth Science (SCIDIP-ES, and** BIG DATA 12-14 Nov 2014 Recommendations, and PV 2015 03-05 Nov Sessions).

### Purge Alert Procedure

As background, Mirko stated that WGISS realizes that it is a responsibility of all organizations holding EO space data to assess the relative value of their holdings and to preserve them for the long term. Sometimes an organization must make the decision to “purge” one or more datasets that could be important to help meeting the mission requirements of another organization.

Data Purge is *to permanently and irrecoverably remove all copies of an EO dataset held in an organization*.

The “Data Purge Alert” procedure aims at preventing, or at least minimizing, the loss of EO space data. Organizations intending for whatever reason to purge an EO dataset should apply the procedure (before purging the data) to inform other organizations with the goal to trigger a possible transfer of preservation responsibility to another interested entity. The procedure will vary slightly depending on dataset uniqueness and purging organization responsibility versus the dataset (e.g. ownership).

Mirko showed a flowchart of the proposed procedure.

The initiator when contacting organizations should provide the following information:

1. Data record description as a minimum in terms of: mission/sensors characteristics, products specifications, temporal coverage, geographical coverage, dataset size, media of storage, archiving format, IPRs, ownership and access policies.
2. List of available associated knowledge (e.g. tools, information)
3. Results and documentation of a Data Record Appraisal (e.g. <http://eros.usgs.gov/government/ratool/>)
4. Overview of constraints and conditions if any
5. Point of Contact

Purge alert response and implementation:

* Organisations contacted in the frame of the Purge Alert procedure (e.g. dataset owners, copy-holders or entities interested to take on the responsibility for the preservation of a dataset) should respond to the “Data Purge Alert” within three months to start conducting negotiations/assessment with the alert initiator.
* CEOS shall be involved when needed through its executive officer (CEO).

Next steps are to consolidate and finalize Purge Alert Procedure, and insert Purge Alert page in new CEOS WGISS website at the top level of WGISS site. The team should also report this information at the next CEOS-SEC meeting and at the CEOS Plenary for information.

Andy suggested that some use cases would be helpful. There was a lot of discussion about where the procedure should be posted. Wyn brought up the topic of datasets at risk because they are not being properly maintained. Not purging, but not preserving.

**Action WGISS-39-14**: Mirko Albani and WGISS Chair to discuss the process of the data purge alert at a near-future CEOS-SEC meeting.

### CEOS Best Practices

Mirko displayed the data stewardship best practices document tree, with Policy documents, technical documents, and the data stewardship best practices/guidelines drafting/approval cycle.

Documents for adoption today are:

1. Preservation Workflow
2. Generic EO Data Set Consolidation Process
3. Persistent Identifiers
4. Purge Alert
5. EO Data Stewardship Definitions (as a contribution to the WGISS Glossary)

Preservation workflow is a procedure recommended to be applied to achieve EO dataset preservation and to optimize their reuse in the long term. Output will be a complete, discoverable, accessible, and useable Earth observation dataset and a series of documents describing the preservation strategy pursued, the implementation plan, and individual activities conducted. Generic workflow should be tailored for individual Earth observation datasets and applicable to heritage, current, and future Earth observation missions. The preservation workflow consists of four phases: Initialization (preservation planning), consolidation, implementation, and operations.

The Generic EO Data Set Consolidation process produces a consistent, consolidated and validated set of "Data Records" and "Associated Knowledge”. It can be applied to any Data Record (e.g. raw data, Level 0 data, higher-level products, browses, auxiliary and ancillary data, calibration and validation data sets, metadata). It should be tailored for each mission depending on objectives, budget availability, operational constraints, and according to the sensor category.

Generic EO dataset consolidation process steps are data collection, cleaning/pre-processing, completeness analysis, and processing/ reprocessing. Mirko showed components of each of these.

The persistent identifier best practice provides recommendations on the use of Persistent Identifiers to Earth Observation mission data, allowing globally unique, unambiguous, and permanent identification of a digital object.

The DSIG proposed the following timeline for future documents:

* EO Data Preservation Guidelines document is under consolidation in order to include comments from LTDP WG, NASA and QA4EO Study, and for alignment with GEO Data Management Principles. The document addresses eight main “themes” consisting of “guiding principles” and a set of “key guidelines” that should be applied to guarantee the preservation of EO space data in the long term ensuring also accessibility and usability. It should be aligned with GEO Data Management Principles.
* Preserved Data Set Content document is under consolidation in order to include comments from LTDP WG, NASA and QA4EO Study and for alignment with GEO Data Management Principles. It provides a description of the composition of the Earth Observation “Preserved Data Set Content (PDSC)” indicating what to preserve in terms of data and associated knowledge and information during all phases of an Earth Observation mission. Mirko listed the principles.

Andy noted that the document tree is a good format for the document elements WGISS is producing. It helps to visualize the work. He added that the WGISS ToR could be written to fit the data management principles. The document tree could also include a link to the associated open source code where it exists.

### Persistent Identifiers Pilot Implementation

Mirko introduced the session with background on persistent identifiers (PI). A persistent identifier is a long-lasting reference to a digital object – a single file or set of files. They are globally unique, unambiguous, and permanent identification of a digital object for locating and accessing it for a long time. These benefit the users by improving data discoverability and accessibility, enabling researchers to cite objects consistently over time, and facilitating re-use of Earth Observation data for new research. They benefit data holders by increasing data visibility, discoverability and usage, allowing tracking of data use and citation, and indicates a level of commitment, and increases the level of trust towards the organisation.

#### **ESA**

Mirko presented the PID implementation test SEASAT, whose scope is testing the implementation of Persistent Identifiers, in particular DOIs, for EO data products at ESA. The elected test case: SEASAT recently reprocessed dataset. OGC metadata standard adopted at ESA comprises two metadata elements that can host Persistent Identifiers: eop:doi - Digital Object Identifier (<http://www.doi.org>), and eop:vendorSpecific (can be used to provide additional attributes in the product metadata). The focus of the pilot test was on PID implementation in the product metadata (both options), ingestion and retrieval from archive. The conclusion is that the eop:doi field should be used for DOIs.

An example of how to get a DOI: DataCite. This is a member of the International DOI Foundation, a global, not-for-profit organisation formed in London on 01/12/2009 aiming to establish easier access to research data on the internet, increase acceptance of research data as legitimate, citable contributions to the scholarly record, and support data archiving that will permit results to be verified and re-purposed for future study. DataCite does not allocate persistent identifiers itself; this is done by its members, who act as allocating agents.

Assigning and registering Persistent Identifiers through DataCite:

* Become a DataCite member; membership is open to all not-for-profit organizations who wish to allocate DOI names and use DataCite as the Registration Agency. Members are eligible to actively take part in the working groups, have full voting rights on all decisions, and may register unlimited DOI names.
* Work with an existing DataCite member; if an organization is only interested in assigning Persistent Identifiers to Earth Observation datasets, then the work should be carried out through a DataCite member. Agreement with a DataCite member is required: <https://www.datacite.org/about-datacite/members>

Components of a PID System (based on DOI) are as follows:

1. The data resource that will receive a persistent identifier (e.g. EO data collection)
2. DOI ‘prefix’, obtained from DataCite: a number that uniquely identifies the providers’ subset of DOIs (i.e. the DOI ‘prefix’ is unique and it is a single number for the data provider/organisation)
3. DOI ‘suffix’: unique for each collection, generated internally by the data provider (e.g. DLR use an online random code generator from http://www.random.org/strings/)
4. Landing page: a web page with information about the data and a download link, hosted by the data provider’s web server
5. An XML metadata file for each DOI, constructed according to the DataCite metadata standard.

To register a DOI, the XML metadata file, the URL for the landing page, and the identifier itself are sent to DataCite, and from there into the DOI resolver system. When a user clicks on a DOI citation, the resolver (http://dx.doi.org/) can then redirect the user to the landing page. The XML metadata will be used for data discovery via online search, metadata-harvesting services, data portals, and data repository catalogues. The DOI and the dataset itself should never change. The data provider must maintain these components: update the landing page on its own web server, update the DataCite metadata by sending an updated XML file, and send new URL to DataCite if landing page location changes.

Next steps at ESA

1. Assess impact and identify necessary steps for PID (DOI) implementation at ESA for EO data holdings according to CEOS Best Practices.
2. Define implementation plan, choose a registration agency, define agreement and sign contract (TBC).
3. Implementation Steps: Create a landing page (use already existing web pages for datasets when available), DOI Metadata generation and DOI integration in the EO-SIP (significant effort), archive system upgrade for query through PIDs (significant effort).

Mirko was asked if they are giving recommendations for what should be included in the landing pages. Mirko replied that it is in the best practices.

#### DLR – Earth Observation Center (EOC)

The DLR - Earth Observation Center (EOC) PID Activities and Pilot Implementation are a contribution to LTDP WG Study - harmonizing approaches. Key points on drafting EOC PID policy are: Central EOC PID Service point-of-contact, DOI selected as PID system to use. DOI generation: random 10-character alphanumeric string + 2-digit checksum (≙ IBAN checksum) 🡪 e.g. mf87c2zjgo59 applied at the collection level to permanent (LTA) data and products only - likely L1 and up (TBD).

DLR has signed contract with DataCite to get an Institutional DOI prefix: 10.15489/ and is developing PID implementation workflow and architecture replacing initial interactive DOI assignment - generation, metadata, landing page.

The DLR PID pilot implementation is the TIMELINE Project, which involves NOAA AVHRR 1 km data consolidation over Europe, and generation of stitched level 1b and 18 level 2/level 3 thematic products. The first DOIs issued to TIMELINE product collections are:

* TL.AVHRR.L1B\_TOA:      10.15489/mf87c2zjgo59
* TL.AVHRR.SYSVAR:        10.15489/hj8s8f57i767

These will be inserted into NetCDF files and archive metadata, will be registered with DataCite and be resolvable, and there will be a manual generation of DOI metadata and landing page.

#### JAXA

Yoshiyuki presented the status of JAXA PID activity. He reported that they have done a study on PID implementations at CEOS agencies, and a study of the impact a PID implementation would have on JAXA systems. In Japan, dataset PID (DOI) discussion officially started in 2014. Japan Link Center (JaLC) was nominated as a DataCite member in Japan. JAXA found that there is increasing interest and implementation within CEOS agencies. CEOS PID Best Practices, and GEO Data Management Principles has clear statements regarding PID.

DOI assignment experiments on scientific datasets is now ongoing in Japan. Nine agencies are participating, including DIAS. Wide open discussion with potential users/data providers was held in February. This revealed two main data providers with different motivations for being part of the dataset PID assignment: Individuals (researchers, scientists), and Data Centers. A policy will be established by September, at the end of the study.

Key points of JAXA’s PID implementation: which PID to adopt (DOI, ARK, URN), which digital object should have the PID assignment, PID string syntax, landing page, RA selection and contract, what needs to be done to existing systems to support PID operation, policy for the long-term commitment, and defining operational procedure with respect to PID.

Preliminary study conclusions:

* DOI for PID is the CEOS Best Practice recommendation; there are no clear advantages of using other (e.g. ARK). JaLC is promoting DOI assignment for datasets.
* Opaque character string for DOI: “Identifier” should have no meaning by itself; this is the CEOS Best Practice recommendation.
* DOI assignment to EO dataset collection is the CEOS Best Practice recommendation, not on browse, ancillary data, or NRT data.
* Landing Page: a good description of the dataset is desired, and navigation for the data access should appear on the page.
* RA selection and contract: DataCite is the number one candidate and is the CEOS Best Practice recommendation.
* What needs to be done to the existing systems to support PID is to prepare landing pages and put in place the “DOI assignment system”, DOI (and metadata) ingestion to RA and maintenance (keeping everything current), and metadata addition to the catalog system. The portal should be enabled to show DOI along with other metadata, show instruction on how to cite the datasets using DOI, and allow the download files to include the above (DOI and citation instruction).
* Policy for the long-term (persistent) commitment. This means annual membership fee to RA, DOI generation and registration, and maintenance of DOI and landing pages. With this commitment in mind, the policy should be explicitly stated in a document, and should be shared among relevant missions/projects.
* Operational procedure with respect to PID should be incorporated in operation manuals, and should be shared among related mission systems.

JAXA findings are: increasing PID use/implementation in CEOS agencies; dataset-DOI gathering momentum in Japan with the leadership of JaLC; DOI implementation requires cost; DOI-included citation should become a firm practice to justify the involvement and find clear benefits (cited-by traceability). Is it the right timing to have PID systems? JAXA is still looking closely at ongoing related activities for the final decision (GEO, CEOS, DIAS, and JaLC).

Michael raised the issue that the DIF is only in English, adding that they do have the facility to put the text in English and French.

#### **NASA**

Andrew Mitchell gave a status update of Digital Object Identifiers (DOI) for NASA EOSDIS data. Andy explained the two options for DOI submission requests, and showed graph of monthly submitted DOIs:

1. Reserve: The Data Provider submits a request to ESDIS to reserve DOIs for EOSDIS products until those products are ready for DOI registration. ESDIS requires that each product be identified with a minimal set of descriptive information (metadata) for initial review. Upon review, ESDIS assigns a DOI with a unique identifier to each product. This DOI is kept in “Reserve” status in the database and on the ESDIS Wiki website until all metadata required for DOI registration is collected. No landing page is necessary to reserve a DOI with ESDIS.

2. Register: The Data Provider submits a request to ESDIS to assign DOIs to EOSDIS data products for registration with EZID\* for public access. Before ESDIS can register DOIs with EZID, all mandatory metadata describing the data product must be present, including the DOI landing page URL. Once the DOI is reviewed by ESDIS for completeness, the DOI is assigned a unique identifier (if not already reserved) and “Registered” with EZID for immediate public access.

Andy also showed a diagram of DOI Workflow for NASA DAACs, and listed DOI Mandatory Metadata for Registration, plus additional DOI information required by ESDIS. He described the DOI process: First assign a unique DOI. Next develop the metadata, develop the landing page, and lastly, register with the service provider. He showed an example of a landing page, with data citations and DOI reference, and the recommendations to EOSDIS DAACs.

Jerome asked who is responsible for the landing page. Andy replied that the data centers are responsible for them. In discussion is whether the agency should take over this responsibility. Jerome wondered how this will work in the long term. Maintaining them is a big task.

For near-real-time data it is not recommended to include the DOIs in the Best Practices. Andy commented that this would be a question for NOAA and EUMETSAT.

**Action WGISS-39-15**: Mirko Albani to discover if NOAA and EUMETSAT are using DOIs for near-real-time data.

### Preservation of Software and Documents at CEOS Agencies: Approaches and Lessons Learned

Mirko gave a presentation on preservation of software and documents. He began with a few definitions, and some challenges:

* Long-term accessibility and exploitability of Earth observation data requires that not only sensed data but also associated knowledge is properly preserved and made accessible.
* Information technology is changing rapidly affecting EO digital data preservation and use: risk of obsolescence of file formats, extant hardware and operating environments that could make data unreadable on the physical and logical level.
* EO data hardware and software management environments are made obsolete by technology evolution in short timespans: need to find ways to preserve both data and their execution context.
* For some digital objects, like software programs, the absence of source code may be a problem. Legal aspects, like copyrights or copy protection mechanisms, can make this even more complex.

Associated knowledge is also important, and he listed elements of software/tools, and information. The features of information format for Digital Preservation include Open Format with no patents or licenses preferred; In wide use and endorsed by other established repositories; variety of writing and rendering tools available for the format; interoperability (tools available for conversion to other formats); inclusion of error-correction capabilities and data integrity features; supporting a stable mechanism for metadata management; and lossless formats preferred for image, video and audio.

Mirko listed examples of information format for digital preservation of text documents, images, metadata, and multimedia, noting that PDF/A format is suitable for documents, and TIFF and FITS is suitable for images.

Mirko elaborated on software and applications preservation since preservation of software and data records are two sides of the same coin. He listed and explained several approaches: Preservation of hardware (techno-centric), emulation (data-centric), virtualization (data-centric), migration (functionality-centric), cultivation (process-centric), hibernation (knowledge-centric), deprecation (budget-centric), and procrastination (I'll-get-round to it eventually -centric).

He added that hardware is also important, and the easiest way to ensure that there will always be hardware on which to run your software is to preserve the hardware (and its operating system and any other reliant software).

### **ESA Approaches to Preservation**

Mirko explained ESA approaches to preservation, beginning with a diagram of the preservation life-cycle. He noted that two categories of documents to be preserved have been identified:

* Mission Specific documentation (mandatory for preservation) for which two preservation formats are currently considered for implementation 🡪 PDF/A and FITS
* Other kind of documentation (e.g. papers, presentations) for which one preservation file format is enough 🡪 PDF/A

Images related to EO missions considered preservation format is TIFF or FITS depending on the level of metadata schema to be preserved. The Metadata files are preserved in XML. The Multimedia file format is under analysis. The Software Preservation approach at ESA is based on different strategies depending on the software (e.g. virtualization, periodic migrations, hibernation).

ESA has found that dependency on legacy hardware, difficult to maintain, was a critical issue for long-term exploitation of the EO products archive and a typical situation for IPFs of historical missions. IPFs and MMFI components can be executed in virtual environments and have been installed on Cloud for many pilot and operational activities.

IPFs: A series of mission data processors was virtualised: AVHRR, SeaWiFS, Landsat TM, Envisat ASAR, Envisat MERIS, Envisat AATSR, Envisat GOMOS, Envisat MIPAS, Envisat SCIAMACHY, Envisat MRA2/MWR, ERS SAR, ERS Scatterometer, ERS GOME, and ERS Altimeter.

MMFI: Migration of MMFI building elements including IPF processors to a virtual-machine hosted environment, identifying and eliminating as much as possible dependencies on specific HW and related SW libraries.

Mirko described two specific examples with SEASAT and JERS-1.

#### **Discussion**

Satoko commented that this is very interesting for WGISS, and would like to discuss further. She proposed to have a dedicated session at WGISS-40.

Mirko listed information formats, asking what format is the most convenient for preserving:

* Text Documents (word, txt, ppt) 🡪 PDF/PDF/A, FITS, other
* Images (bmp, tif, jpg, gif) 🡪 TIFF, FITS, other
* Metadata file 🡪 XML, ASCII
* Video (avi, vob, m4v, mov, mpx, etc.) 🡪 MJ2, other

One of the members commented that many of the documents are currently only on paper. They can be converted to PDF, but cannot be edited. An editable file would be very helpful.

Richard said that for CNES PDF-A is the preferred format for documents, though they encounter problems with some tools’ used for PDF-A conversion. He suggested asking the planetary community, as they have been concerned with this for a very long time. Andy suggested consulting ESIP. Yonsook said it would be great to know what USGS and NCDC are doing.

Gabor mentioned that there is a special working group dedicated to long term preservation of historical maps.

Andy said missions are required to provide the software, but primarily so users can understand how products were generated; it is not expected to be executable. All new missions are required to preserve based on the new preservation standards.

Satoko said that JAXA may have to upgrade software every 3-5 years to extend the software life without modification, and do not know how major the upgrades will be. They are still wondering in which direction to go.

Richard said at CNES certain data that has existed for a long period of time is reprocessed every two years, so that keeps the software fresh. For the data, CNES has a team that verifies that the data is still up to date, and if not, reformat and reprocess to standard formats. For packed bit data they do a formal description of the data files and are able to describe and read the data. CNES is all data. When the data achieves phase F – the data are frozen and all the information is gathered, ensuring formal descriptions. The level of work depends on the mission. He mentioned a couple of missions for which they have done the exercise. In some extreme cases it took several months to even find the documents. And some data not documented at all.

USGS is a leader on software and document, and data associated knowledge preservation. They could provide significant input.

**Action WGISS-39-16**: Andy Mitchell to reach out to the NASA working groups on data stewardship for possible presentations at WGISS-40.

**Action WGISS-39-17**: Mirko Albani to plan a session for WGISS-40 on document and software preservation.

## CEOS Interoperability Interest Group

The CEOS Interoperability group consists of the IDN, FedEO, CWIC, and OpenSearch.

### International Directory Network

Michael Morahan gave a report on the activities of the IDN Interest Group. He began with the response to the IDN actions from WGISS-38:

Action WGISS-38-9:*Yves Coene (ESA) to check whether the SKOS description of satellite and missions is in the IDN*. Response: GCMD/IDN opened up the KMS SKOS API (dynamic search) without authentication.

Action WGISS-38-10:*Andy Mitchell (NASA) to review the use of the term DOI in the IDN DIF metadata field to determine if it is appropriate to use PI instead.* Response: GCMD/IDN has updated the DIF (version 10) schema to use the appropriate Persistent Identifier.

#### Upgraded IDN Site Demonstration

Michael gave a demonstration of the new IDN site, which has been updated to incorporate current look/feel of the GCMD search to improve navigation. It is possible to search by keyword or by free text.

Participants provided feedback, and Michael asked them to review the site and notify them of any issues or questions.

**Action WGISS-39-18**: IDN Interest Group to reword the phrase “create CEOS dataset” on the new IDN website by WGISS-40.

**Action WGISS-39-19**: IDN Interest Group to identify alternate terminology for the term “portals” used on the IDN website by WGISS-40.

#### What to Include

Michael clarified that the GCMD has all holdings, whereas the IDN only has CEOS member holdings.

Not all datasets in the IDN are coming from CEOS members. A clear definition of what belongs in the IDN is needed. Yonsook suggested that if the CEOS member is either producer or distributor of the data, it could be included. You have to have the affiliated datasets to make it useful. In CWIC they note the producer and the distributor; both identifiers provide key information.

Recommendation for the IDN is to include datasets of CEOS Agencies, CEOS Agency Associates, and datasets sponsored by CEOS.

Recommendation is to keep the IDN is as and report the metrics to CEOS based only on CEOS holdings.

The CEOS agency members that have portals in the IDN are ESA, JAXA, EOSDIS, and NOAA. Associate members are GOFC, UN, and CEOP. CEOS Other Agencies that have CEOS data are WWF (World Water Forum), AMD (Antarctic Master Directory), and ANTABIF (Antarctic Biodiversity Forum).

Recommendation: remove portals that have been incorporated in the past, but are not CEOS affiliated: GISD, Human Health and Disease Portal, Human Health and Disease Portal Services. Also remove CEOP.

Continuity of support for the IDN includes CEOS/GEOSS services such as IDN portals, CWIC (records, members, QA, metadata mappings), discovery of CWIC records (OpenSearch or CSW servers), OpenSearch server, and CSW server for GEOSS services. IDN can also deploy GeossDataCore tags on request from agencies.

Next steps: Maintain continuity of DIF format/content, support DIF 9 for IDN partners, DIF format evolution (In parallel with UMM-C), and work with IDN partners to ensure high quality DIF content. IDN records that would eventually move to the CMR are transition IDN (CEOS) metadata records into the CMR (by the end of the year), and support QA of CEOS and other non-NASA metadata in the CMR.

#### GCMD Keyword Status and Plan

Michael reported the release (March 26) of Keyword Version 8.1 (Land Surface and Atmosphere). Changes are new/updated Land Surface and Atmosphere keywords (246 new keywords and 10 updated keywords). Version 8.2 (Atmosphere, Ecosystem, and Terrestrial Hydrosphere) proposes to submit to NASA ESDIS standards office (ESO) for review by October 2015, and release by March 2016. Version 8.3 (Water Vapor, Water Quality/Chemistry, and Ecosystems) proposes to submit to ESO for review by May 2016, and release by October 2016. Further Releases are expected to follow a six-month cycle.

#### Unified Metadata Model – Collections (UMM-C) Upgraded

This is used by the NASA EOSDIS community as a guide during metadata generation for the Common Metadata Repository (CMR). It takes into account existing collection metadata formats (DIF, ECHO, and ISO 19115-2).

The DIF-10 was developed to support for UMM-C compliance; it includes additional required fields, new fields, and enumerations for specific fields. Michael detailed the DIF-10 changes, and gave the following benefits of the changes.

* UMM-C compliance for the Common Metadata Repository allows easy of metadata record ingest into the new repository
* Developed with ISO in mind; help maps ISO fields to the DIF
* New Fields to describe the datasets: AdditionalAttribute, ProductLevelId, Version.
* Restructured existing fields to better describe the datasets: Platform > Instrument > Sensor hierarchy, PersistentIdentifier, and Spatial\_Coverage.

The DIF-10 Translator converts DIF-9 to DIF-10, DIF-10 to DIF-9, ECHO-10 to DIF-10, and fills in missing required UMM-C fields where possible. It is implemented in Adapter Framework (not XSLT conversion), and supports file-based “drop box” capability (GUI in the works).

#### IDN Metrics

Michael gave metrics for the IDN site for March 2014 – April 2015, and showed diagrams by continent and country. Similarly for the production CSW Service. He graphed the continuing rising trend of IDN metadata records, and number of IDN records updated. He also displayed number of CWIC datasets by topic, the DIF breakdown of the GCMD, and US GEO/GEOSS metrics.

Michael reminded that the IDN statistics page is a good source for CEOS DIF counts by parameters, by Source\_Name Bucket (Platform Type), by IDN\_Node, and by Data Center.

Michael mentioned the following resources:

* docBuilder: News and downtime notifications on metadata authoring tool ([ceos-idn-docbuilder@lists.nasa.gov](mailto:ceos-idn-docbuilder@lists.nasa.gov))
* Interoperability Forum: Release announcements and proposed changes to DIF and SERF ([ceos-idn-interop@lists.nasa.gov](mailto:ceos-idn-interop@lists.nasa.gov))

Wyn asked for difference between the GCMD and CMR. Michael replied that the CMR, at a high level, will be a cleanup and merging of ECHO and GCMD.

### CEOS WGISS Integrated Catalog (CWIC)

Yonsook Enloe and Ken McDonald opened the CWIC session.

#### CWIC Status

Yonsook Enloe gave a status report on the project, noting that EUMETSAT and ISRO (MOSDAC) became operational data partners; ISRO will start working to integrate NRSC to CWIC. The annual CWIC meeting in February focused on CEOS OS BP, improving searching, additional monitoring and metrics needed, and support for client and data partners.

The team also produced CEOS OpenSearch BP Developer’s Guide, and reported that CWIC and IDN conform to CEOS OpenSearch BP. The team is discussing how to support more complex organizational hierarchy via IDN and CWIC. Yonsook reported CWIC/IDN synchronization through mapping tables, and updating of the Data Partner Guides (CSW & OpenSearch) and Client Partner Guides (CSW & OpenSearch). Exception handling and status codes to accommodate the diversity of data providers and the data provider limitations for search and results have been implemented, better monitoring of data partners, and improved metrics.

CWIC OpenSearch updates include production of the CEOS OpenSearch BP Developer’s Guide, support for StartIndex and for dc:identifier, CWIC OSDD updates for (space/time) data center limitations on searches, completed alignment of CWIC implementation with CEOS OS BP, and updated the Data Partner Guide and Client Partner Guide.

In the context of improving search relevancy, Chris Lynnes reported that “free text keywords collection searches can return many non-relevant search results. Just because users would rather SEARCH WITH free text keywords, that does not mean we have to FIND BY free text methods. Implementing good relevancy ranking may be the single most important thing we could do to improve our search.” Return dataset in the most relevant order, given keywords provided by the user; likely user intent, inferred from empirical experience, type of user, referrer, and keyword. See additional details in Chris’ slides in February Team meeting webpage. <https://wiki.earthdata.nasa.gov/display/CWIC/2015+CWIC+Annual+Meeting>

Yonsook noted that the data producer is a different organization than the data distributor. For example: GHRSST - NOAA archives the GHRSST data collections from many organizations. The “data producer” name should be noted to be different than “data distributor”. Also, there can be multiple data centers for one organization – indicating the structure will require additional information to be in the DIF or to be harvested from the data provider or kept in CWIC internal tables. OS results do not provide this.

CWIC/IDN Synchronization is implemented through IDN “Tagging”. Tags (“CWIC”) are kept outside the DIF metadata record. Can search IDN for only those DIFs with specific tag. One example is this VC example: LSI team tags data collections with “LSI” tag; LSI wants to search for only LSI tagged data; LSI Portal can perform collection search at IDN for the “LSI” tagged data collections plus search criteria (keywords, space, time); after use selects data collection of interest, can search CWIC.

In terms of monitoring and metrics, Yonsook said that they need to improve the information about scheduled and unscheduled data partner outages. To this end they will start an email list for cwic-ops for partners to send info; planning an area of CWIC webpage to post information about scheduled downtime and known outages.

CWIC data partners provide a major source of satellite data inventory search and data access in GEO. Current data partners include NASA, USGS (LSI), INPE, NODC/GHRSST, CCMEO, EUMETSAT, and ISRO (dev partners: AOE, NOAA CLASS, ROSCOSMOS, ISRO 2nd data center):

* NOAA data centers consolidated into one virtual data center.
* NASA working to make all EO unrestricted online satellite data accessible from CWIC
* GHRSST (SST VC) is working to make all their data accessible from CWIC at NOAA NCEI/NODC. NODC adding data to this CWIC connection.
* CCMEO provides search and email data ordering to access to Radarsat-1 & Radarsat-2
* EUMETSAT became operational with CWIC in March 2015
* ISRO (MOSDAC) became operational with CWIC in May 2015. Will start working to integrate NRSC.
* AOE working to add CRESDA (HJ1A, HJ1B, CBERS01, CBERS02, CBERS2B), Beijing-1 (BJ-1), and NSMC/Feng-Yun (FY3A, FY2D, FY2E).
* ROSCOSMOS to making progress on their CWIC connection

Yonsook listed the CWIC data partner support that is provided for existing and for new partners.

Yonsook concluded saying that both the IDN and CWIC integrated with GEO, and accessed by the GEO components (Geo Web Portal, DAB, and GENESI). IDN provides collection metadata for over 28,000+ data collections available to GEOSS (May 2015), with 12,000+ 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 about 1900 (May 2015 figure) data collections with inventories (about 70 million granules, adding granules daily from live missions). Providing technical support for GEO component teams accessing IDN and CWIC.

Yonsook displayed the WGISS/CWIC webpage, outlining all the documents and resources available.

#### CEOS OpenSearch Developer Guide

Doug Newman\* reported that the CEOS OpenSearch is achieving a robust and interoperable implementation. He began giving the rationale for two separate documents:

* CEOS Best Practices: how to implement an Open Search implementation for an earth science data provider
* CEOS Developer Guide: client and server focus, discusses how to how to link two providers together via two-step, and is a handbook for CWICSmart, IDN and CWIC system

Originally the development guide was a set of guidelines for interoperability of IDN and CWIC Open Search APIs, and programmatic construction of a user interface to the IDN and CWIC OpenSearch APIs. It is now a set of guidelines for general Earth data Open Search API interoperability, enabling the use of CWICSmart as a client to your Open Search API, and client and server development for Open Search. It is built on top of the CEOS Open Search Best Practices document.

In detail, the developers’ guide conforms to the Open Search Specification, to the relevancy and parameter extensions, to the OGC extensions. It takes pertinent ideas from ESIP best practices (Spatial MBR extents for all). It also includes WGISS recommendations to make the best attempt at coercing clients to provide client ids, to extend ESIP ‘two step’ idea to link IDN dataset -> CWIC granule searching, to embrace hypermedia – HATEOAS, for relevance, for utilizing the parameter extension for better clarity of service, for temporal (Dublin Core) extent representation, and for propagation of parameters from collection to granule resource.

Doug showed a diagram showing the search process and the two-step search process. He also displayed CWIC Smart as an example: Any Open Search implementation that conforms to the CEOS Developer Guide can be queried using CwicSmart. He showed a diagram of a CWIC Smart dataset search and the CWIC Smart granule search. He noted that using other APIs, the user can use CWICSmart as a search tool for their inventory. He showed a diagram showing dataset compliance of OSDD, ATOM, free text, bounding box, temporal end, CEOS DG, CEOS BP, and two-step.

In conclusion, Doug recommended coding the OS service to the CEOS Developer Guide to gain a ready-made client, and a simple gateway into the CWIC system; and coding the OS client to the CEOS Developer Guide to gain a gateway into the IDN/CWIC aggregation of providers, but a lot more compliance boxes need to be checked.

Richard suggested that the name of the guide should be CWIC Developer Guide. Doug said it began as a CWIC developer guide, but was opened up to provide a better user experience. The document is much broader than CWIC. It can be a user interface for any OpenSearch implementation. CWIC Smart name is a legacy name.

**Action WGISS-39-20**: OpenSearch Interest Group to follow the WGISS procedure for releasing the Developer’s Guide (similar to what was done for the CEOS OpenSearch Best Practices document). Final comments should be due by June 15.

#### CWIC Partners

##### ISRO Connector

Yonsook reported for Nitant Dube, giving the status of ISRO. She reported that ISRO is an operational CWIC data partner. Data for 17 DIFs from ISRO Meteorological and Oceanographic Satellite (MOSDAC) Data Centre integrated with CWIC including capability for Direct Ordering.

The following Satellite and Metadata/Products are available from MOSDAC: INSAT-3D IMAGER, INSAT-3D Sounder, SARAL, Megha-tropiques, INSAT-3A VHRR, INSAT-3A CCD, and Kalpana-1 VHRR.

Future plans include testing and fine-tuning MOSDAC connector for issues if any, registration of remaining DIFS for MOSDAC Connector, and integration of the NRSC connector.

##### CAS Academy of Opto-Electronics (AOE)

Guangyu Liu reported progress on Catalogue System Integration of China EO Data. First he reviewed the architecture diagram showing their connector to CWIC. Three data centers in China are integrated into the CWIC system through China connector of CWIC:

-National Satellite Meteorological Center (NSMC): FY3A, FY2D, and FY2E

-BeiJing-1: Beijing-1\_Multispectral

-China Center for Resource Satellite Data and Application (CRESDA): ZY01, ZY02, ZY02B, HJ1A, and HJ1B

AOE Responsibilities are running and maintaining China connector system, upgrading and optimizing China connector system, and coordinating with CWIC project about activities related to sharing satellite data. Guangyu Liu displayed Diagram of OpenSearch interface architecture. Adding OpenSearch protocol to their middleware. Present progress includes a study of the standard of OpenSearch interface, completion of the design of OpenSearch implementation, and the OpenSearch service of Chinese node is going to be provided this year.

Ordering process for FY series satellite data: satellite data will be provided by email. They are working with Dr. Kang to add the email information to CWIC system.

Yonsook commented that WGISS is really looking forward to connecting to their data.

##### ROSCOSMOS

Tamara Ganina reported on the Roscosmos Geoportal as data provider connection scheme with CWIC connected to the Roscosmos Geoportal native search service. Currently service specification search on the Common Catalogue cannot directly connect to CWIC directory. Therefore, in this case on the side of the directory CWIC realized adapter for connection of the service. Thus, metadata products with Russian satellites contained in the Common Catalog will be made available to foreign users via directory CWIC.

According to test results within the framework of CWIC tasks Roscosmos is now releasing a new version of the search service (it takes into account the issues that have arisen during testing).

Main activities:

1. Tests of the Roscosmos Geoportal native search service were completed by the CWIC team, the Roscosmos Geoportal search service has been updated according to tests results (bugs, performance, error processing capabilities, and extended response parameters). The test server and new documentation release is almost ready.
2. The main logic of CWIC-Roscosmos connector was implemented by the CWIC team.
3. Platform and instrument descriptions for registration in IDN test server were prepared. It should be submitted to IDN production server.
4. The submission of the data collections is pending.

Tamara listed the platforms and instruments for IDN registration. She added that they prepared the description of the satellites and camera devices in xml format for registration them in IDN. They upload them to the test IDN server. For publication this information on the primary server IDN to reconcile this list with the Roscosmos list. They have also prepared the description of the data collection.Each data collection corresponds to data from particular instrument. Description collection in xml format suitable for IDN. For publication on the primary IDN server agreement with Roscosmos will be necessary.

Data collections for IDN registration: data collections contain satellite images. Each data collection corresponds to data from a particular instrument.

Next step is obtaining government approval.

##### CWICSmart

Yonsook introduced this saying that they are working toward making this fully open source. Doug Newman\* began the presentation explain what CWIC-Smart is. It is a client for IDN and CWIC Open Search, for any CEOS Open Search Developer Guide compliant implementation, for any CEOS Open Search Best Practices compliant implementation, and for any Open Search specification compliant implementation.

For dataset search the application uses the OSDD to dynamically render an html form. Using the template and the parameter extensions it can be determined whether a parameter is mandatory, its type, its bounds and whether it is a ‘core’ parameter like searchTerms. The user submits the form and the results are presented. It uses HATEOAS to generate links for traversing through the results. But in the absence of such links it uses elements mandated by the base OS specification. Search links within the entries to provide ‘Search this <resource>’ links that perform two-step searches.

For granule search it parses 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). All shared parameters are propagated to this search form. From the OSDDs it is known whether they are ‘the same’. The user submits the form and the results are presented. It also uses HATEOAS to generate links for traversing through the results, icon links within the entries to render browse thumbnails to the user, and ‘enclosure’ links within the entries to allow the user to download a resource (granule in this case).

CWICSmart code is on the Earthdata Code Collaboratory, and will be made public at some point soon in the future. It is a simple Ruby-On-Rails application, and instructions are available to replace default assets with custom ones.

In conclusion, Doug stated that CWICSmart can be used to query any Open Search implementation with varying degrees of user experience:

* BEST – CEOS Developer Guide compliant
* Better – CEOS Best Practices compliant
* Good – OGC extension compliant
* Fair – Open Search specification compliant

CWICSmart will soon be open-sourced to allow providers to re-badge the application to their needs, and can be found at: <https://api.echo.nasa.gov/cwic-smart>

Yoshiyuki asked about GCMD compliance; Doug said it is only waiting for the DC identifier. He also asked if CWIC and ECHO support the DC identifier. It does.

Once the source code is open source, the agency can put their own logo on it, and can constrain the access as desired. It is a basic excellent client that can be built on.

#### CWIC via Browsers

Archie Warnock began saying that not all data providers are accessible via OpenSearch. CWIC provides an OpenSearch (and CSW) gateways to remote data partners, regardless of their native search protocol. Custom clients like CWIC-Smart use this capability to greatly simplify and enhance the process. But browsers can do OpenSearch, too.

It is the essence of the two-step search: Search collections at IDN, which provides pointers to CWIC granule search and a template URL to search CWIC. CWIC OpenSearch provides a template URL to search remote data partners. Clients (ideally, fully-programmatic ones) can do this for you. But browsers can do this too.

The process to search the IDN OSDD is, for example, for ISRO INSAT-3D Imager L2B SST.

Search IDN OSDD:

1. Request generic IDN OSDD (it tells you how to search the IDN)

<http://gcmd.gsfc.nasa.gov/KeywordSearch/default/openSearch.jsp?clientId=CWIC>

2. Find the IDN OSDD URL template in the Atom response

3. Use the URL template to request the dataset OSDD: [http://gcmd.gsfc.nasa.gov/KeywordSearch/OpenSearch.do?searchTerms=INSAT\*&MetadataType=0&output=atom&Portal=cwic&clientId=fromgcmd](http://gcmd.gsfc.nasa.gov/KeywordSearch/OpenSearch.do?searchTerms=INSAT*&MetadataType=0&output=atom&Portal=cwic&clientId=fromgcmd).

4. Find the CWIC OSDD URL template in the Atom response.

Search CWIC OSDD (Use CWIC OSDD for INSAT-3D Imager L2B SST)

1. Request the OSDD for the dataset (it tells you how to search)

<http://cwic.wgiss.ceos.org/opensearch/datasets/3DIMG_L2B_SST/osdd.xml?clientId=fromgcmd>

2. Find the template in the Atom response

Search CWIC for granules (Search CWIC for ISRO INSAT-3D Imager L2B SST granules)

1. Get the request URL (it does the search)

<http://cwicdev.wgiss.ceos.org/opensearch/granules.atom?datasetId=3DIMG_L2B_SST&startPage=1&count=1&timeStart=2013-10-01T00:00:00Z&timeEnd=2015-05-04T23:59:59Z&geoBox=0.843296,-81.04153,163.15671,81.04153&clientId=fromgcmd>

2. Look for granules in the Atom response

In conclusion: OpenSearch is simple; it is easy to implement (for programmers – it is just HTTP). It is functional, even without a specialized client. Users gain access to remote collections that do not natively support OpenSearch by going through CWIC.

#### CWIC Service Quality Monitoring and Reporting

Lingjun Kang\* gave a presentation on CWIC service quality monitoring and reporting, whose purpose is to monitor the performance of CWIC request/response handling, monitor the status of data provider’s metadata repository, monitor metrics of client, and incident alert and issue tracking. Its implementation includes metrics/log collecting module, remote server status monitoring script, and metrics/log analyzer.

Eugene Yu\* gave a demonstration of query interface metrics, daily visit trend, metrics by data provider, and client metrics.

#### GEO Community Portals

Ken McDonald gave a presentation on the GEO community portals. He began with a brief report on the GEO Work Plan Symposium. This includes a review of progress of current tasks (infrastructure, institution and development, and information for societal benefits tasks). It also includes a discussion of proposed GEO Strategic Plan 2016-2025: Implementing GEOSS - Areas of Action and Implementation Mechanisms. AT the symposium they will also propose tasks to accomplish the goals of the Strategic Plan.

The GEO Infrastructure Implementation: the focus of the Infrastructure Implementation Board (IIB) is the development of the GCI components and associated processes. Observation/Concern from the WPS are limited utilization of GEOSS and the GEOSS Common Infrastructure (GCI) by the communities it is intended to support, and need to focus more on data access. Two existing mechanisms are in place to address the concern: Architecture Implementation Pilots, and Community Portal Activity.

Ken displayed a diagram of the GCI role in GEOSS. GCI key components are the Component and Services Registry (CSR), the Discovery and Access Broker (DAB), and the GEOSS Portal. Clearinghouse functions are fully provided by DAB. Under development is the revised component and services registration process, and the user registration process. Missing in the diagram are contributions of community components.

As good and necessary as the GEOSS Portal is, communities still wish to develop their own interface to relevant data and services that provides visibility to their particular program, supports integration of EO and non-EO data, and enables integration of specialized community services (e.g. helper applications, data processing, sensor webs, model webs…) Ideally Community Portals will utilize GCI for discovery and access of GEOSS resources. The GCI can enable CPs to present a community-specific view of GEOSS data, services, standards, etc. Communities will need assistance, and for this there are GCI documentation, recommendations, best practices, tutorials, etc. Ken displayed a diagram showing the community portal role in GEOSS.

The Community Portal activity is evolving in scope to look at range of possible interactions between GCI and Community Infrastructures, to allow GCI to better leverage community capabilities, and to promote closer engagement of user communities in sustaining and enhancing GEOSS. The Diagram of community systems role in GEOSS shows the potential interfaces between communities and the GCI. WGISS can be considered one of the communities.

The Community Portal Recommendations document has a section on outreach, and another for instruction.

Going forward, Community Portal Activity needs closer engagement with community activities, and begin to assemble recommendations based on community experience.

Ken directed the following questions to WGISS (specifically CWIC and FedEO):

* Does WGISS have the same concerns with respect to community engagement?
* Should WGISS more actively engage with GCI, GEO SBAs?
* Is a WGISS-Smart client needed or is DAB sufficient?

The community systems role in GEOSS: if you go through the GEOSS portal you can access CWIC and FedEO through the DAB. This is a system that brokers the brokers. There are interesting aspects to these potential implementations.

Yonsook recommended getting this information to the VCs – enough infrastructure is in place so they can access what they want. Satoko is currently leading the water portal which is one of the community portals, and it has already interfaced with GCI DAB. WGISS can write some improvement requests to them. WGISS can work with the GCI DAB team as a community portal. Ken said that dialog across multiple communities will be very valuable to the GCI developers.

### 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 mission data based on HMA standard interfaces. He described the operational, test and development environments, and explained operational setup. In alignment with CEOS and GEO approaches for OpenSearch, FedEO is based on two-step search. Mirko mentioned that future metadata handling in FedEO will be collection search. He described the FedEO OpenSearch implementation and showed some statistics. FedEO now has a page on the WGISS website. Some guides related to FedEO (client partner, data partner) are available from the website.

Mirko listed ongoing and future activities, which include consolidation of FedEO concept/scenarios as part of the ESA and European EO Ground Segments. Consolidation of interfaces with existing European catalogues and addition of new catalogues, e.g.; population of FedEO Collection Metadata Catalogue for ESA and European Missions; consolidation of access from GEO DAB Broker, tests with non-European catalogues; and collaboration/interoperability/interfacing with CWIC, NASA ECHO.

FedEO uses RESTful interfaces for discovery and extends HATEOAS for ordering. FedEO also uses RESTful interface for ordering and for interface for linked data. Mirko showed a diagram of the linked data model. The have achieved improved search response using faceted search. Mirko also displayed a diagram showing FedEO collection search and granule search, and mentioned the SmartHMA Android client, which is Open Source.

Yves Coen gave a live demonstration of FedEO.

### CEOS OpenSearch Project

Yoshiyuki Kudo introduced the CEOS OpenSearch Project session.

The purpose of the OpenSearch project is to promote the use of the OpenSearch standard as a means of data discovery for Earth Data providers, and to define the expectations and requirements of candidate OpenSearch implementations. It also aims to remove ambiguity in implementation where possible, facilitate the aggregation of results between disparate Earth Data providers via OpenSearch common standards, allow 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.

Yoshiyuki reported that the OSDD Version 1.0 is complete as of today. He gave an outline of the document and listed the 17 best practices, which are based on OpenSearch implementations at CEOS agencies. He explained details of a few of the Best Practices.

#### Discussion on V1.0 Finalization

One last remaining item to be resolved is CEOS-BP-011: *Support of dc:identifier returnable and the corresponding geo:uid queryable to allow for “search-by-ID”*; should it be a “Requirement” or “Strongly recommended”? He listed the comments and use cases so far.

Jerome felt it was best to keep it as a requirement. Doug said strongly recommended (he got some push-back from ESIP), wondering what is gained by making it mandatory. Andy was concerned that ESIP will come up with their own best practices. But Yoshiyuki reminded that the two step search was an idea that came from ESIP.

Yves said it is a unique, not a universal identifier. Mirko said if all the agencies around the table are in agreement with this, then it should be a requirement. Satoko said that JAXA is not yet compliant but plans to be soon; it is important to know what to do, and would prefer that if the item is useful and works well that it be a requirement. Otherwise future systems may not work as well as expected.

The final decision was to make it a requirement. There is no need to re-do the comment period for the document.

#### Project Conclusion

Yoshiyuki said that with the finalization of Version 1.0 of OSDD, the CEOS OpenSearch Project is successfully concluded. In addition, they have achieved exposure and spreading to relevant agencies and public. But it was agreed that the project web page as well as mailing list should be kept active.

Richard suggested that the CWIC Developer’s Guide be a contribution of the OpenSearch Project. The guide was developed by CWIC, but if FedEO could provide comments and perhaps some input it could be a more generalized contribution from WGISS. Richard’s proposal is to not close the project but to ingest the document into the project’s activities and prepare it as a contribution. This would include sending it out for comments, and have the CEOS OpenSearch Best Practices, and the Developer’s guide together from the OpenSearch Project.

Richard said that the people that were in this project group should be the ones to circulate it and modify it, but Satoko said she needs to find out from the point of view of the JAXA contract. Yoshiyuki said he could do the work. Richard is only recommending a quick second round on reviewing the document; send the email next week, give one month.

At end of June there should be a clear picture of the comments on the Developer’s guide. In mid-September WGISS can report to SIT that the BP document is complete, and finalize the Developer’s guide at WGISS-40. Both documents can be on the web page for OpenSearch (and CWIC and FedEO can also link it).

**Action WGISS-39-21**: Richard Moreno to prepare a two-minute video to present at the CEOS Plenary demonstrating discovery and access using FedEO and CWIC.

# WGISS Plenary Session, Part II

## **WGISS-39 Actions**

Michelle Piepgrass reported that all actions from WGISS-38 are closed, and listed the actions resulting for WGISS-39.

**Action WGISS-39-01**: WISP to talk to SEO about transferring the WGISS email distribution lists to CEOS and the WGISS documents to the CEOS Document Management System.

**Action WGISS-39-02**: Mirko Albani, Richard Moreno, Andy Mitchell and Wyn Cudlip to select a team that will review GEO Task IN-02-C1: Advances in Life-cycle Data Management, and make recommendations looking toward the GEO 2016 Work Plan.

**Action WGISS-39-03**: Richard Moreno and Andy Mitchell to approach Jonathon Ross about taking on the role of WGISS Collaboration Facilitator, then approach the SIT and the SEC for adoption.

**Action WGISS-39-04**: Michelle Piepgrass to include a session at WGISS-40 for deciding the future of the OpenSearch Project.

**Action WGISS-39-05**: Michelle Piepgrass to work with WISP to implement the recommended changes to the WGISS website, including the short WGISS “purpose/description” paragraph.

**Action WGISS-39-06**: IG/P leads to implement recommended changes to their pages on the WGISS website.

**Action WGISS-39-07**: Michelle Piepgrass to work with WISP to migrate WGISS documents to the CEOS document management system; also discuss with the SEO the wording on the CEOS documents landing page.

**Action WGISS-39-08**: Andy Mitchell to consider a session for WGISS-40 on the retirement of FTP.

**Action WGISS-39-09**: Mirko Albani to prepare a presentation for WGISS-40 on ESA’s cloud processing.

**Action WGISS-39-10**: Technology Exploration Interest Group to plan a session on Open Source Software practices inviting each agency to present at WGISS-40.

**Action WGISS-39-11**: WGISS Exec to respond to SEO’s request to conduct a data format study to analyze the advantages and disadvantages of using GEOTIFF, NetCDF, and HDF formats in a Data Cube. Are there computational and data assimilation advantages? The SEO/GA can provide WGISS with a small data cube for testing.

**Action WGISS-39-12**: WGISS Exec to respond to SEO’s request to develop an automated data ingestion scheme that utilizes CWIC/FedEO to ingest new data products into the Data Cube infrastructure as functional layers. The SEO is working with USGS to create automated scripting and requests for Landsat data. Can this work for the Sentinel missions? No other missions have open access or would be desired as “regular” data layers.

**Action WGISS-39-13**: WGISS Exec to respond to SEO’s request to conduct a data compression study to compare different compression approaches (e.g., JPEG-2000, HDF-5) for the Data Cube storage. Consider initial Data Cube formation, delivery to countries and deployment, Data Cube re-gridding, and regular operations of algorithms.

**Action WGISS-39-14**: Mirko Albani and WGISS Chair to discuss the process of the data purge alert at a near-future CEOS-SEC meeting.

**Action WGISS-39-15**: Mirko Albani to discover if NOAA and EUMETSAT are using DOIs for near-real-time data.

**Action WGISS-39-16**: Andy Mitchell to reach out to the NASA working groups on data stewardship for possible presentations at WGISS-40.

**Action WGISS-39-17**: Mirko Albani to plan a session for WGISS-40 on document and software preservation.

**Action WGISS-39-18**: IDN Interest Group to reword the phrase “create CEOS dataset” on the new IDN website by WGISS-40.

**Action WGISS-39-19**: IDN Interest Group to identify alternate terminology for the term “portals” used on the IDN website by WGISS-40.

**Action WGISS-39-20**: OpenSearch Interest Group to follow the WGISS procedure for releasing the Developer’s Guide (similar to what was done for the CEOS OpenSearch Best Practices document). Final comments should be due by June 15.

**Action WGISS-39-21**: Richard Moreno to prepare a two-minute video to present at the CEOS Plenary demonstrating discovery and access using FedEO and CWIC.

## Adjourn

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

# Glossary of Acronyms

API Application Programming Interface

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

COTS Commercial Off-the-Shelf

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

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