MINUTES

OF THE

41st MEETING

OF THE

CEOS WORKING GROUP ON
INFORMATION SYSTEMS AND SERVICES

(WGISS)

Canberra, Australia

14 March to 18 March, 2016

Hosted by

Geoscience Australia

(GA)

Table of Contents

[1 WGISS Plenary Session, Part I 5](#_Toc450205351)

[1.1 Welcome, Introductions, Adoption of Agenda 5](#_Toc450205352)

[1.2 Host Welcome and Logistics Information 5](#_Toc450205353)

[1.3 Host Opening Address 5](#_Toc450205354)

[1.4 WISP Report 5](#_Toc450205355)

[1.5 WGISS Chair Report 5](#_Toc450205356)

[1.6 Review of WGISS-40 Actions 7](#_Toc450205357)

[1.7 Discussion on Joint Activities with WGCV 7](#_Toc450205358)

[2 Data Preservation 8](#_Toc450205359)

[2.1 Data Stewardship Interest Group Update 8](#_Toc450205360)

[2.2 Preservation of Associated Knowledge 8](#_Toc450205361)

[2.3 ESA Maturity Matrix – Model for Harmonization 8](#_Toc450205362)

[2.4 ESA-NASA Joint Effort for Historical Data Recovery 9](#_Toc450205363)

[2.5 Landsat Global Archive Consolidation 9](#_Toc450205364)

[2.6 Thesaurus 9](#_Toc450205365)

[2.7 Persistent Identifiers 10](#_Toc450205366)

[3 Data Access 11](#_Toc450205367)

[3.1 CEOS Water Portal Project 11](#_Toc450205368)

[3.2 CEOS OpenSearch Project II 11](#_Toc450205369)

[3.3 International Directory Network (IDN) 11](#_Toc450205370)

[3.4 Federated Earth Observation (FedEO) 12](#_Toc450205371)

[3.5 CEOS WGISS Integrated Catalog (CWIC) 12](#_Toc450205372)

[3.5.1 CWIC Report 12](#_Toc450205373)

[3.5.2 ISRO Report 12](#_Toc450205374)

[3.5.3 ROSCOSMOS Report 12](#_Toc450205375)

[3.5.4 NOAA Report 13](#_Toc450205376)

[3.6 OpenSearch Validation 13](#_Toc450205377)

[3.7 WGISS Connected Data Assets 13](#_Toc450205378)

[3.8 Access Discussion 13](#_Toc450205379)

[4 Data Use 15](#_Toc450205380)

[4.1 COVE Tool 15](#_Toc450205381)

[4.2 CEOS Data Cube Prototypes in Kenya and Colombia 15](#_Toc450205382)

[4.3 Future Architecture for Data 15](#_Toc450205383)

[4.4 Recovery Observatory Project 16](#_Toc450205384)

[5 Technology Exploration 17](#_Toc450205385)

[5.1 Semantics 17](#_Toc450205386)

[5.2 Big Data 17](#_Toc450205387)

[5.2.1 Bridging the Gap between HPC and High Performance Data Analysis 17](#_Toc450205388)

[5.2.2 Recent Work on Remote Sensing Cloud Computing and Big Data Processing 18](#_Toc450205389)

[5.2.3 NoSQL Influenced Solution for EO Data Exploration 18](#_Toc450205390)

[5.2.4 Discrete Global Grid Systems: A New Way to Manage 'Big Earth Data’ 18](#_Toc450205391)

[5.2.5 Storage Structure and Efficient File Access 19](#_Toc450205392)

[5.2.6 Bringing the Processing Close to the Data 19](#_Toc450205393)

[5.3 Technology Exploration Discussion: Semantics and Big Data 19](#_Toc450205394)

[6 WGISS/WGCV Joint Plenary 21](#_Toc450205395)

[6.1 Welcome and Introductions 21](#_Toc450205396)

[6.2 GA Welcome Address 21](#_Toc450205397)

[6.3 CSIRO Welcome Address 21](#_Toc450205398)

[6.4 CEOS Executive Officer Report 21](#_Toc450205399)

[6.5 GEO Secretariat Report 22](#_Toc450205400)

[6.6 GEO Work Plan IN-02-C1 / GEO Strategic Plan GD-05 23](#_Toc450205401)

[6.7 Introduction on WGCV/WGISS Joint Efforts for Future Data Access 23](#_Toc450205402)

[6.8 Carbon Action Items 24](#_Toc450205403)

[6.9 Discussion on Joint Interaction 24](#_Toc450205404)

[7 Agency and Liaison Reports 25](#_Toc450205405)

[7.1 Geoscience Australia (GA) and Commonwealth Scientific and Industrial Research Organisation (CSIRO) 25](#_Toc450205406)

[7.2 Global Spatial Data Infrastructure (GSDI) Association 25](#_Toc450205407)

[7.3 Indian Space Research Organization (ISRO) 26](#_Toc450205408)

[7.4 Japan Aerospace Exploration Agency (JAXA) 26](#_Toc450205409)

[7.5 National Aeronautics and Space Administration (NASA) 26](#_Toc450205410)

[7.6 National Oceanic and Atmospheric Administration (NOAA) 26](#_Toc450205411)

[7.7 US Geological Survey (USGS) 27](#_Toc450205412)

[8 WGISS Plenary, Part II 28](#_Toc450205413)

[8.1 Future Meetings 28](#_Toc450205414)

[8.2 Chair Summary 28](#_Toc450205415)

[8.3 WGISS-41 Actions 28](#_Toc450205416)

[8.4 Adjourn 30](#_Toc450205417)

[9 Glossary of Acronyms 31](#_Toc450205418)

List of Participants

Institute for Remote Sensing &

Digital Earth, CAS Lizhe Wang

CNES Richard Moreno

CSIRO Robert Woodcock (co-Chair CEOS FDA team), Matthew Steventon, Peter Wang, Matt Paget, Arnold Dekker

ESA Mirko Albani (WGISS vice-Chair), Andrea Della Vecchia\*, Philippe Mougnaud\*, Damiano Guerrucci\*

GEO Osamu Ochiai\*

Geoscience Australia Chris Pigram, Stuart Minchin, Jonathon Ross (CEOS-DCEO), Simon Oliver, Matthew Purss, Joshua Sixsmith, Leo Lymburner, Jeremy Hooke

GSDI/HUNAGI Gábor Remetey-Fülöpp

ISRO Nitant Dube\*

JAXA Satoko Miura

National Computing Infrastructure Lesley Wyborn, Rui Yang, Jingbo Wang, Joseph Antony, Pablo Larrondo

NASA Andrew Mitchell (WGISS Chair), Brian Killough (CEOS-SEO), Yonsook Enloe, Eugene Yu\*

 Michael Morahan, Doug Newman\*, Michelle Piepgrass (WGISS Secretary)

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

ROSCOSMOS Tamara Ganina\*

USGS Kristi Kline, Thomas Cecere (LSI, co-chair CEOS FDA team)

WGCV Albrecht von Bargen (Chair)

\* Via web conference or email

# WGISS Plenary Session, Part I

## Welcome, Introductions, Adoption of Agenda

Andrew (Andy) Mitchell (WGISS Chair) welcomed the participants to WGISS-41. Andy thanked Geoscience Australia for all the excellent arrangements, and asked those present to introduce themselves. He reviewed the agenda and it was adopted with no modifications.

## Host Welcome and Logistics Information

Jonathon (Jono) Ross welcomed the WGISS-41 participants to Canberra and to the Geoscience Australia (GA) facility; he described the logistics of the meeting for Wi-Fi, lunch, breaks, group photo, evening activities, and places of interest in Canberra.

Jono highlighted that GA is Australia’s national geoscience agency, with the goal of applyinggeoscience to Australia’s most important challenges. Focus areas include resources wealth, community safety, water resources, maritime jurisdiction, fundamental geographic information, and national geoscience capability. GA integrates Australia’s geological and geophysical survey, Australia’s land information and mapping group, and Australia’s centre for remote sensing.

## Host Opening Address

Dr. Stuart Minchin, Chief of Division of Environmental Geoscience, GA, welcomed the participants to the GA facility. He said that GA is very happy to host both CEOS working groups, and especially glad for the positive relationship between GA and CEOS. Two of their top concerns this year are the CEOS Chair Initiative and the data cube (DC) efforts.

Stuart noted the high level of interest in the DC, adding that the work has been done in a very open and deliberate way; the team is small, so they appreciate contributions from the CEOS community. Having the CEOS community behind the effort is important due to its global impact. The concept is not new (other communities have been doing this for years); data continues to be managed as it is captured, but with the DC the archive becomes the instrument.

Australia is also very interested in the Copernicus hub, and is intending to develop the Sentinel data for Australia and Oceania; there is a lot of interest in the data, but the data is large, and it must be structured in a way that can be analyzed.

A third priority is the effort of the OGC. The DC is structured in tiles, with Landsat pixels within tiles: a nested grid concept. GA has invested effort in co-chairing with the OGC to develop standards. Incompatible geologies result in huge efforts in moving from one projection to another, and with geo-processing tasks. A discrete global system can make the project quite simple, but expertise and a gridding standard could be the basis of developing a sustainable global system.

CEOS is a critical component of GEO, and the relationship with GEO is very important to Australia. Realizing the vision of GEO is near; a federated version with big data systems does not work, but further development of nodes where the data is located is an exciting opportunity.

Stuart wished the participants good luck and an enjoyable stay in Canberra, and developing pleasant memories of Canberra and Australia. Andy thanked him, and complimented him on the nice facility, and very pleasant location.

Andy asked if their Sentinel hub includes all instruments. Stuart that yes, it does, but they are only hosting the data that is relevant to Australia and to the region. Andy referred to the discussion at CEOS of the benefits of the data hubs; it will be very beneficial to the region for Australia to be part of this; Stuart mentioned that China is also keen to participate.

Andy also noted that the OGC global grid system is very interesting to WGISS, and that will be the next step of the DC and ARD work. With OGC working on this, the huge barrier (collect, calibrate, mosaic, map, etc.) will be lowered. Andy asked if they plan to put it in an OGC test-bed. Stuart replied that work is progressing along that path.

Tom Cecere commented that the earlier paradigm was a country’s data set, geospatially aligned, free and open; but the new one is to integrate the data and harmonize it to get the power of the data across time, space and spectral region. Stuart agreed this is critical development that needs to occur, with the gradual lowering of the previous policies, contributing to the global pool of data and that with China, Australia and Oceania, the conversations are happening. Stuart agreed that it is really important to highlight the issues to CEOS and GEO, especially when decisions are being made in the absence of technical knowledge.

## WISP Report

Martin Yapur provided details on the WGISS Infrastructure Services Project (WISP) team members, contact information, WGISS and CEOS structure. He also described recent changes in the WGISS website, which contains the latest work and structure of WGISS, and is designed to better communicate WGISS functions. He also provided instructions for participating in web conferencing and for submitting presentations

**Action WGISS-41-15:** WGISS Exec to work on the organization of the Documents web page and content of documents.

## WGISS Chair Report

Andy reported on recent activities of WGISS and CEOS issues related to WGISS. He began with the WGISS organizational structure, which is redesigned to be aligned to data.

Andy discussed the CEOS Work Plan 2016-18, and pointed out three new tasks in 2016 that impact WGISS:

* OUT-7: Data Cube infrastructure development; WGISS will support the SEO in the development of a general CEOS Data Cube infrastructure.
* CARB-9: GFOI Data Services Pilot Projects for Kenya and Colombia; WGISS will support the SEO in the delivery of Data Services Pilot Projects (based on the Data Cube architecture) for Kenya and Colombia.
* VC-25 Increase the visibility of land surface imaging data holdings; WGISS will work in conjunction with the LSI-VC to ensure relevant datasets are visible through WGISS Interoperable Standards.

Andy listed four pre-existing tasks:

* DATA-2: Full representation of CEOS Agency datasets in the IDN and accessible via CWIC/OpenSearch. WGISS began discussions with the Australian centers (Geosciences Australia / CSIRO / Bureau of Meteorology /Australian National University and National Computational Infrastructure) in order to get their data accessible via WGISS interoperable standards (i.e. IDN, CWIC).
* DATA-7: Study on future data architectures; The WGISS Future Data Architecture (FDA) subgroup is participating in the CEOS Future Data Architectures Study Team. WGISS will host a 1-2 day(s) "Future Data Architecture" workshop during WGISS 42 in Sept 2016. WGISS is contributing to Section #5 ‘Review of relevant CEOS and agency activities and ambitions’ (WGISS efforts on Big Data and Cloud Computing).
* VC-1:  List of Relevant Datasets from VCs; WGISS has not received any updated list from the VCs. This will be discussed during SIT 31.
* CARB-08-35 – The CEOS Carbon Subgroup (recommended in Carbon-Action-38) will develop guidelines for appropriate data use of satellite data and data products. This will require improved interaction between the carbon cycle community and the satellite community; comprehensive review of the current use of data products, including current data limitations; and reconciliation of methodological differences and spatial compatibility. Such interactions may include co-sponsorship of joint workshops targeting specific data needs and investment in community product assessments, especially for key in comparison exercises. WGISS is identified as lead for this action although it does not evenly fit within WGISS duties/expertise. The CEOS SIT Chair Stephen Briggs suggested that WGISS offer to host a Carbon Workshop for the CEOS Carbon Subgroup. Andy mentioned that guidance from Jono is needed for CARB-08-35 as it does not clearly tie to the work of WGISS. Jono agreed, but added that WGISS does have knowledge of the structure of the data, and that is what Stephen Briggs might have in mind. Andy agreed to discuss this at the SIT meeting.

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

A discussion followed about combining GD-2 and GD-7, though it would be helpful to know the reasons they were kept separate. Andy said the DMTF and Data Sharing subtasks would be good ones to combine. Jono asked that the suggestions be formalized if he is to go to CEOS and GEO with suggestions.

Andy also reported that the SIT Technical Workshop is in April in Frascati. He mentioned Session 9, ‘Copernicus Sentinel Data Uptake’ which will be a report on the latest ground and space infrastructure developments, and is an opportunity for CEOS VCs/WGs/initiatives with an interest in exploitation of Sentinel data to report on their experience and any issues and recommendations to optimise access and exploitation of the data. Andy suggested writing up what we have from the WGISS-40 workshop. Jono suggested that the session is an opportunity to talk about what is happening more broadly with agencies outside Europe. There will also be a Future Data Architectures side meeting, and an LSI-VC session, but it is at the same time as the GEOGLAM session.

Andy suggested several changes for the semi-annual WGISS meetings:

* A dedicated discussion time has been added to the WGISS meeting agenda for many sessions to allow appropriate time for feedback.
* WGISS Exec is considering a proposal to shorten the five day meeting to a three-day high level (Plenary) meeting, and a two-day workshop dedicated to a particular area (Sentinel, data access formats, cloud computing). These workshops would provide the opportunity to advertise and invite others to participate, even from outside the CEOS community. Many agreed this would be useful, and that the output could be presented as a white paper at SIT meetings. Points that were raised are that it would require a lot of planning, the key experts need to be identified, taking advantage of the location of the meeting, and the participant list needs to be considered carefully so that no one is left out. It is easier to justify a meeting if there is a technical workshop, and Jono suggested that another option is to have one WGISS meeting per year, and one WGISS workshop. Another option is to have a workshop on Future Data Architecture, and what ARD means for remote sensing data in September at ESRIN. Robert Woodcock said one of the impacts of the DC was a change of format from tiff to netcdf. Issues will come up that will be of use to the broader community and September would be good timing. Robert noted that by September the FDA team would have something to report; the key technical concepts will be identified by July to begin identifying what it takes to bring the experts in.
* Joint WGISS meetings with other CEOS WGs/VCs/ad Hoc groups are being considered, as well as a dedicated WGISS side meetings at CEOS SIT and Plenary meetings to attract Chair/vice-Chair of these groups.
* Proposal to initiate a series of monthly webinars for various Technology Exploration topics. Satoko Miura is in full agreement and has a list of possible topics.

Andy discussed WGISS support and outreach. He noted that in the MIM there is a field for mission links where the data can be accessed; it is intended to make data more discoverable. Alignment of IDN/MIM keywords and CEOS EO Handbook and the Data policy portal is important; some links exist from the MIM to the data policy/access portal but some links are missing and the context (search criteria) are not passed in the linkage.

In order to initiate dialog with other WGs, VCs, Ad-hoc groups, Andy recommended requesting one-on-one meetings with VC/WG/Ad-hoc team members to assess their needs with regards to WGISS capabilities. This might be easier to do at SIT meetings and SIT technical workshops. Andy has PoCs for WGCV, WGClimate, WGCapD, WGDisasters, GEOGLAM, GFOI, LSI VC, ACC, Ocean Color Radiometry, Precipitation, SST, Ocean Surface Topography, and Ocean Surface Vector Winds; he invited suggestions of others. Mirko said he can easily contact WGClimate and WGDisasters. The list should be expanded, with names of the PoC.

Tom suggested that having a pick-list of WGISS activities, where WGISS has had successes and where WGISS can contribute. Other outreach activities may be developing something similar to the “Faces of CEOS” video for WGISS activities. One video could be produced for each focus area. These can also be pointed to WGCapD and the VCs. The WGISS Interoperability video shown at 2015 CEOS Plenary was well-received. This video demonstrates the search and access of satellite data across multiple CEOS agencies by making use of a common set of commonly defined WGISS interoperability standards. The link for the video is <https://youtu.be/xWkuDxGdRjs>

**Action WGISS-41-08: In order to discover how WGISS can assess the needs of other CEOS working groups, virtual constellations, and ad-hoc groups with regards to WGISS capabilities:**

**Action WGISS-41-08a: WGISS members to identify, from within their own agencies, members of said groups.**

**Action WGISS-41-08b: WGISS-Exec to coordinate one-on-one interactions with said contacts.**

**In order to improve WGISS outreach:**

**Action WGISS-41-09a: WGISS-Exec to investigate ways to develop additional outreach materials.**

**Action WGISS-41-09b: WGISS Exec to start planning personal engagement for invitations to future meetings.**

**Action WGISS-41-09c: WGISS Exec to put together a list of the benefits and capabilities of WGISS to assist the conversation. This list will also be put on the WGISS website.**

## Review of WGISS-40 Actions

Andy reviewed the actions from WGISS-40, discussing the outcomes of each. All are closed with the exception of those moved to WGISS-41.

## Discussion on Joint Activities with WGCV

Andy presented slides from WGISS-37 on good candidates for cooperation with WGCV. The following topics were discussed:

* Quality data metadata fields for the key sensors and products, including spatial and spectral. The characterization may be even more important with ARD. Kristi Kline suggested finding ways to segment the issue.
* Data access of CEOS test site information and cal/val sites; there are 15 DIFs for these in the IDN at this time.
* Linking of CEOS tools and accessibility of tools.

For the carbon action items, Mirko said it is important to include something of data quality; perhaps WGCV are able to provide input.

For near real-time (NRT), it was clarified that the first version released (NRT) is not validated, and only minimal quality information is provided; the second level is validated after two weeks. In fact, it can take up to 20 days before it is science-ready.

For reprocessing of data, communication is important: when it is planned, what the changes are. Users should be notified, and the approach should be predictable. It would be very useful to have a methodology for just CEOS agency data. Kristi added the additional point of how long to keep the starting version of data being reprocessed. What is the user feedback from replacing older collections with newer ones? Some of the versioning issues go away if the generation pipeline is automated (re-compute rather than store the data). The information needed would be in the metadata. Current practices of re-processing, and best practices of reprocessing would be helpful. Tiers of NRT vs science-ready data could be:

* Orbit information
* BP and current practices of re-processing
* Golden months
* Search relevancy

Richard Moreno raised the point of a landing page of the DOI - a persistent page where a user can find out about re-processing, actions of the spacecraft. If the information can be found and have a uniform format, the PI page would be a good one for the landing page. In France they have done something for the landing page that is quite mature. At NASA they also have product landing pages. DOI landing pages are very short. Effort is underway for the new metadata repository to have a generator to automatically produce the landing page.

# Data Preservation

## Data Stewardship Interest Group Update

Mirko Albani chaired the Data Stewardship Interest Group (DSIG) session. He reviewed the background, purpose and scope of the group. He also summarized past and ongoing activities.

Mirko reviewed some points of the GEO Work Programme and the Data Management Principles Task Force (DMP-TF) that impact the DSIG. He highlighted the Foundational Task GD-07: GCI Development (which includes development of data management guidelines). He gave the general description of the task, and listed the five sub-tasks, highlighting GD-07.03: Collect and analyze the survey results supporting the Data Management Guidelines and develop a process to implement the guidelines for providers. Mirko and Richard are members of the GD-07.03 task team. The full description of the tasks can be found at www.GEO.org. Mirko showed the key milestones of GEO.

Lesley Wyborn inquired about the distinction between the DMP issued by GEO and the DMP issued by USGEO. The USGEO DMPs were submitted to GEO, so should be the same.

DMP Implementation Guidelines are under review and will be submitted for adoption at GEO-XIII in November. An online survey is underway and Mirko is soliciting feedback/responses. Harmonized feedback will be provided by ESA agencies in the frame of Long Term Data Preservation (LTDP) working group activities.

Mirko explained that the GEO governing board oversees tasks and activities, and that the workplan symposium is where the tasks and activities progress are presented. For technical discussions WGISS needs to participate in the GD-07 team; teleconferences occur every three weeks. The GD-07.05 sub-task only has one person.

Mirko also mentioned upcoming conferences: ‘Big Data from Space’, ‘Living Planet Symposium 2016’, ‘iPRES 2016’, and ‘PV 2017 Conference’.

Mirko displayed the DSIG Document Tree, showing policy and technical documents. The Glossary of Acronyms and Terms submitted by the Long Term Preservation of EO Space Data document is in good shape and can be issued, with the expectation that updates be made as comments are provided; it is a working document. Andy made the recommendation to expand the Glossary to include the Glossary of Science Keywords. The document was officially adopted.

The PI Best Practices document was updated with recommendations and use case scenarios for PID assignment to NRT products. Andy asked if PIDs include documents and instruments. Mirko replied that so far only data (datasets, and data collections) are included. PIDs to documents makes a lot of sense, but not so much for instruments. The document was officially adopted, and it was agreed that it should be kept as stable as possible.

The Purge Alert process was accepted by CEOS; it is being circulated and will have a two-month review period. It was decided to put the Purge Alert web page on the WGISS website under the Preservation node, with a news banner on CEOS and WGISS pages. The SEO may have additional suggestions on how to publicize it.

Mirko displayed a proposed timeline for future documents, including Software and Documents Preservation Best Practices.

**Action WGISS-41-12:** Andy and Martin to check to see if the USGEO Data Management Principles information is feeding to GEO.

## Preservation of Associated Knowledge

Mirko explained that associated knowledge includes software, tools and information (documents, images, metadata, multi-media, etc.) At previous meetings, WGISS has had discussions on collection of information and software and document preservation aspects, and presented experiences and lessons learned. WGISS still needs a harmonized approach for EO data associated-knowledge preservation.

Mirko displayed the table of contents of the Best Practice document. Mirko asked if anything was missing, and Andy suggested that the sustainment of associated-knowledge should also be included. For example, for documents there is no guarantee that formats will persist. Mirko showed an example of the format recommendation (text as: PDF/A, FITS; images as: TIFFS, FITS; metadata as: SML, ASCII; multimedia as: MPEG-4, MJ2). FITS seems to be a stable standard format, and ESA is doing a pilot to see what the effort is to transform documents to FITS. Lesley asked about practices at a more international level, and said she could provide him with some contacts. Mirko noted that the idea is to have a practical method to provide to EO agencies. These are examples of recommendations that might change; the team will address and investigate what others are doing.

For software and tools, the concept is very complex. The idea is to focus on the type of mission, and its relevance, coverage, size, storage media, and archiving. Software preservation steps are to preserve, retrieve, reconstruct, and replay, and the important attributes are functionality, environment, dependencies, architecture, and user interactions.

Three specific cases should be considered: future missions, historical missions, and current missions; hardware limitations should be eliminated. Andy suggested including open source software with future missions. Mirko showed tentative milestones for the document.

Robert asked if they are covering reproducibility as well, since different hardware may produce different results. It may be worthwhile preserving the documentation on how the data was processed, in case backtracking is needed; in some cases the hardware could be preserved as well. He added that the knowledge itself is useful, even if it cannot be reproduced.

## ESA Maturity Matrix – Model for Harmonization

Mirko Albani reported on the ESA Maturity Matrix Model for Harmonization. This matrix can be used as a scoreboard and a roadmap for improvement, and also can provide data quality and usability information. The matrix can help determine the maturity level which can be helpful to decision-makers. He showed a diagram of the maturity matrix. The nine components are preservability, accessibility, usability, production sustainability, data quality assurance, data quality control/monitoring, data quality assessment, transparency/traceability, and data integrity. These are graded on five maturity states.

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

An exercise has been performed to verify the compatibility of the Data Stewardship Maturity Matrix with respect to GEO Data Management Principles and guidelines. Each DMP has been mapped into Maturity Matrix Components and the rating obtained through DMP implementation has been derived. This exercise has allowed identification of possible areas of improvement for the Data Management Principles. The Data Stewardship Maturity Matrix is highly compatible with GEO DMP, and possible areas of improvement for the Data Management Principles have been identified.

Mirko proposed as next steps the collection of comments and suggestions from WGISS members on the maturity matrix, consolidation of CEOS EO Data Stewardship Maturity Matrix, and its inclusion in the WGISS Stewardship process (when, how). Mirko solicited and received confirmation from WGISS that this is an interesting approach.

Andy remarked that he liked the matrix concept and can see the need in other areas, like data access and use. AGU is developing a matrix also for data use. Rob added that GA is developing a provenance matrix. These tools really help providing a good tool for decision makers. Mirko wondered if a maturity matrix would also be helpful for ARD. Rob replied that the matrix gets really complex when looking at more than one instrument type; it helps to break the problem down, and understand the costs associated with each. The maturity matrix provides insight into moving from one point to the next, whereas the checklist does not.

## ESA-NASA Joint Effort for Historical Data Recovery

Mirko discussed the ESA-NASA Joint Effort for Historical Data Recovery project. The datasets to be addressed in the coming months are Nimbus-7, Adeos-1, MOS-1, MOS-1b, SPOT-1, and SPOT-2. The ESA “LTDP+” (2017-21) Programme will support the recovery, preservation, access and valorization of heritage data of European interest. Examples of some recovery activities that have been done include ERS-1/2 Level 0, and SEASAT SAR.

For NASA, Andy described NASA’s Making Earth Science Data Records (ESDRs) for Use in Research Environments (MEaSUREs) program, which will support the data of NASA interest. As two examples, he mentioned a unified and coherent land surface temperature emissivity, and a Greenland ice mapping project.

Discussion has begun between ESA and NASA, as part of their bilateral cooperation activities, for the identification and recovery of high priority datasets starting from NIMBUS-7, SEASAT, and HCMM. The objective is to recover and possibly harmonize the identified data sets (when complementary data holdings are available) for uniqueness and completeness analysis, format alignment, and reprocessing campaigns. Some of the challenges are that data are very scattered; approximately 50% are unique, and data quality is good but temporal coverage is scattered. Existing tools cannot be used to open ESA Nimbus CZCS data holdings.

Planned activities are cooperation with NASA to understand how the ESA CZCS data holdings can be improved and to discuss a possible harmonization/alignment of the NASA and ESA datasets in order to achieve a consistent, consolidated Nimbus-7 CZCS dataset. New possible datasets have been identified, with mission, sensor, owner and dataset location.

Lesley commented that Dave Gallagher led a project for recovering Nimbus data. It is difficult to prioritize the datasets to be rescued, so it is nice when the scientists actually give direction on where to start. Even as the satellite arm of GEO, all agencies have some non-satellite data to consider. Lesley and Mirko will take the opportunity to exchange information.

## Landsat Global Archive Consolidation

Kristi Kline presented the USGS Landsat Global Archive Consolidation (LGAC) program. She listed the objectives of the project, major accomplishments, and status of ongoing activities.

Remaining work is to improve ingested data that is not capable of producing L1Ts (a Landsat correction process), continue to refine process for reading data off of old media, address parts and equipment concerns for old drives, and develop MSS data format converters. A key factor to the project is having good technical support on hand that understand the machines when getting into trouble.

## Thesaurus

Richard Moreno gave a presentation on thesauruses within the context of EO data. A thesaurus is a dictionary in the form of organized list of terms, with a notion of hierarchy and association; it also may give the origin of the terms, languages available for the same term, and synonyms. The main utility is to share the same vocabulary between different entities, interoperability, and to group a set of metadata by theme.

The CNES thesaurus is in English, and organized in four main themes: Earth, space, life, and condensed-matter sciences. It is represented as a tree, and available in several formats. Richard displayed a diagram showing the relationship between all thesaurus formats. He also listed other thesaurus (NASA, EUROVOC, and UNESCO), and specific thesaurus: IDN, PDS, GEMET, IVOA. CNES has the intention to evolve the thesaurus. The work is very useful for CNES, but may have usefulness perhaps at the CEOS and GEO levels.

Lesley commented that the Research Data Alliance has a vocabulary services interest group with very experienced members. It is probably operating at a higher level, but a lot WGISS may be able to leverage from it.

Richard wondered if thesaurus is an area for collaboration with WGISS. Andy agreed that it is definitely needed and would like to see this done alongside the work already done by the IDN. A lot of agencies have used these keywords as a basis for their ontologies; it is a natural extension of the glossary of terms.

Yonsook commented that the IDN had the advantage that it is operational and funded. NASA is making it easier to submit or change keywords, and the GCMD keywords to describe all NASA’s products. Lesley commented that it is essential to be governed by a vocabulary and should be implemented in the early stages.

Richard suggested an analysis at the technical level even though the most probable solution is the IDN. Andy added that more community buy-in so is needed, and that the feedback from CNES’ thematic method can be leverage for the IDN. Michael noted that the CNES group has been very interactive with the IDN team and proposed keywords to be added.

Thesaurus is something that is interesting to WGISS; Richard will start on a small document about it with the help of Veronique.

**Action WGISS-41-06c: Michael Morahan, Richard Moreno to begin a new initiative on a thesaurus (EO keywords).**

## Persistent Identifiers

Richard Moreno gave a presentation on new input for CEOS persistent identifier best practices. He began with a comparison between LTDP and RDA, and listed the RDA recommendations for DOI. The results of the LTDP WG analysis are as follows: The Research Data Alliance (RDA) is composed of different WGs and in particular the aforesaid RDA recommendations were delivered by Scalable Dynamic Data Citation Working Group under the Data Citation WG. The RDA recommendation gives a set of 14 clear rules that make dynamic data citable. The approach is aimed at dynamic data, where any part of the dataset could change at any time; to give a PID to a data set like that, some kind of timestamp is needed to make sure to the desired version. The Earth Observation time series data sets can be dynamic, but in a very limited way. If anything within the existing time frame changes, it would need a new version and a new PID. The CEOS best practices are really for larger-scale data archives.

Going forward, no changes are needed for the WGISS PID Best Practices, but there should be discussion with the Data Citation WG members in order to explain the CEOS context and recommendations. After a conversation with Data Citation WG, some minor updates could be evaluated. These include re-wording of some recommendations, and developing a new use case: if truly dynamic data sets need to be managed, where any part of the data set could potentially change at any time, consider the RDA time-stamped query store approach.

Richard stated that WGISS should explore the difference between DOI and PI and ARK. It seems that DOI is more suitable for scientific data, and ARK for documents and libraries. Andy said this is the main reason NASA went to PI. This could be a good presentation topic, and ties to getting metrics. Richard noted that CNES is thinking to change from ARK to DOI. The team will organize a teleconference with RDA to discuss these topics.

**Action WGISS-41-05: Richard and Mirko to organize telecon with RDA on the best practice and recommendations for DOI/PI.**

# Data Access

## CEOS Water Portal Project

Satoko Miura introduced the CEOS Water Portal, which is meant to be a one-stop service for the water community around the world. Although only a few use cases were registered, the portal has more than 10 data partners for model, in-situ, and observation data. The portal uses OpenSearch, OpenDAP, and CSW. Now that the portal is complete, its operation will be transferred to Tokyo University in April. It is very likely that the portal will remain unchanged.

Satoko reviewed the concept, history and architecture of the project and noted the following lessons learned: WGISS is comprised of data/service providers, engineers and system developers, not users, so collecting user requirements/feedback is important and challenging. To overcome the challenges, an appropriate PoC must be obtained and a sufficient preparation period for adding data partners must be scheduled, since internal coordination within each data partner agency may take a long time. Using standards for the interface with data partners greatly assists the implementation. Continuous improvement for the user interface should be planned, and should persist through several years.

Andy suggested that it would be helpful to have a discussion (session or teleconference) on the points raised. At NASA they have gone so far as to hire an outside vendor to investigate requirements directly with the users, and this approach has worked well. They have tried different approaches to survey the users, taking the developers out of the picture. Requesting feedback is not as effective, and querying the user community can be very costly, though they use user surveys in the evolution of systems.

Jono commented that metrics of the number of citations of the data is important, adding that difficulties arise as contacts change roles. Jono asked about the Goal diagram that was displayed, noting that registration of a use case can be difficult without any incentive. Curation of the portal is part of the cost, and is the responsibility of the portal curators. Outreach and training people to try new capabilities is essential.

It was agreed that the Water Portal Project page should be disabled after a few months, but the discovery of the portal should remain.

## CEOS OpenSearch Project II

Mirko gave the presentation of the CEOS OpenSearch Project II. He began with an overview of the goals of WGISS to establish OpenSearch Best Practices of the CEOS catalogue systems. He also reviewed the reasons behind developing an OpenSearch Best Practice. He noted that the CEOS OpenSearch II best practice project began at WIGG-40 with the following targets:

1. Consolidation and finalization of updated CEOS OpenSearch Best Practice document by WGISS 41 (Action WGISS-40-11)
2. Updating of the CEOS Developer Guide document in accordance with the BP finalisation
3. Start discussion about CEOS Open Search Conformance Tests document

At WGISS-40 a number of recommendations for the Developer Guide and Best Practice documents were proposed; in agreement with the WGISS-40 recommendation and conclusions, the two documents have been kept separate, as follows:

Best Practice to collect and describe requirements about OpenSearch interface, trying to be as much implementation/technology independent as possible. This significantly reduces the needs to maintain and update the document.

Developer Guide to provide implementation guidelines, permitting an OpenSearch search engine to be aligned with Best Practice requirements. The document might need more frequent maintenance due to evolution of technologies.

A subset of Developer Guide requirements were incorporated in the Best Practice, either fully or partially. There are still some open issues listed in the BP that need to be properly addressed. Next steps are to finalize the Best Practice and to begin the Developer Guide.

Yonsook noted that Doug Newman that the team did a great job of moving things to the BP, and wondered if the developers guide may not be needed as most of the content is in the BP. However, Andy and Mirko preferred having a formal BP document that is fixed, and a developer’s guide that is adaptable. Even if the Developer’s Guide is short, it is better to keep it since WGISS already reported to the CEOS Plenary that there would be two documents. The Developer’s Guide can help the developer to implement, and can also contain lessons learned. Merging into a single document would have a major impact on the proposed roadmap (in addition to frequent reissuing). Yonsook noted there are core requirements and others are optional, and Jerome Gasperi will confirm that the minimum requirements are sufficient for interoperability. The participants confirmed having a common vision about the Best Practice’s and Developer Guide’s roles, and the decision to keep them separate.

Mirko reported that the OpenSearch Best Practice v1.1D4 was released at WGISS-41, and is available for final review cycle; open issues (listed) need to be properly addressed, and corrective action agreed. It will be finalized by the end of June. CEOS OpenSearch Developer Guide will be updated in accordance with a common vision of best practice.

## International Directory Network (IDN)

Michael Morahan gave a presentation on advances and metrics of the IDN. He discussed the DIF-10, which has a richer number of fields than DIF-9, and is restructured to better describe the datasets. The IDN team has provided a translator from DIF-9 to DIF-10, and Michael explained how this translator works. The docBuilder, for import and export of metadata to and from the IDN, is also adapted for DIF-10, and allows for user feedback. The DIF-10 and docBuilder resources were listed.

Mirko highlighted the need for a common metadata model, and suggested that CEOS develop one to meet the various levels of requirements for discovery. Andy volunteered to give a presentation on what NASA is doing with the UMM to standardize their metadata model. The Data Stewardship Interest Group is the appropriate group to consider this. Andy added that he has contacts at NASA that would also be helpful.

Michael discussed the objectives of the ThirdGeneration Search Interface which include building upon the current search interface with new functionality, and designed to increase search precision and/or recall.

Michael gave metrics of CEOS DIF counts, CEOS DIFs by Parameter, CEOS DIFs by Platform Type, CEOS DIFs by IDN\_Node, and CEOS DIFs by Data Center, showing consistent increased usage.

Martin asked if NASA is contemplating continuing the services of the IDN for the foreseeable future. Andy replied that it will continue as a contribution from NASA to CEOS. He added that the reorganization of the metadata within NASA was meant for higher efficiencies within NASA.

**Action WGISS-41-12c: Andy Mitchell to give a presentation at WGISS-42 on what NASA is doing with the UMM to standardize the metadata model.**

**Action WGISS-41-12d: DSIG to give a presentation on the investigation of a CEOS-recommended metadata model.**

## Federated Earth Observation (FedEO)

Mirko Albani presented an update of FedEO. He reported that there has been an increase in the number of accessible datasets, and he listed the accessible collections from European agencies, NASA, and JAXA. They are working toward harmonizing to ISO 19139.

The Sentinels scientific data hub improvements were displayed with additional collections as well. Mirko explained how granule discovery and download is done. DLR 158 collections with two-step search are now available through FedEO, and effort with CNES is underway to make the Theia and Pleiades collections discoverable. Other efforts include adding collections for the ESA Online Dissemination, which has a large number of collections, and some improvements with Cryosat-2 that can be accessed through FedEO.

Ongoing activities include an improved collection catalog with automatic discovery of available organizations, platforms and collections.

Service monitoring includes online tools for the administrator to monitor service back-ends, and service information in dataset series search responses (Atom) to be exploited by the FedEO clients. Automated metrics includes generation of spreadsheet information for publication on the CEOS website.

Andy commented that WGISS needs to move the search relevancy topic higher on the list of topics for Technology Exploration, along with ranking of search results. A lot of the organizations are carrying the same data, and it is essential to remove the duplication in the metrics, or to filter the duplication. On the WGISS side it might be done manually, but on the agency side it is going to be a huge problem.

## CEOS WGISS Integrated Catalog (CWIC)

### CWIC Report

Yonsook Enloe gave a report of the status of the CWIC project. She gave an update on the CWIC data partners, and introduced Jingbo Wang who is the contact for GA and CSIRO on their CWIC implementation; the work with Australia is just getting started. She also noted that CWIC supports OGC CSW 2.0.2, CEOS OpenSearch, and the recommendation for NASA ES data systems.

Yonsook reported that NASA is developing access to CWIC from the Earth Data Search Client. The source code for this client will be made open source. The IDN/GCMD has a new implementation; all existing APIs are supported. IDN tagging of data collections (e.g. CWIC, GHRSST) offers seamless access to an identified subset of data. The annual CWIC team meeting was in February; the focus of the meeting was the new IDN implementation and CWIC synchronization with the new implementation. All presentations from the meeting are available online. The CWIC project page has updated metrics. Yonsook added that a system level team has emerged for CWIC development and support.

Yonsook also described the Earth Data Search Client, which will provide access to CWIC very soon. The GCMD keywords are fully integrated into the client and can be used to narrow down the collections. The client can perform collection search and granule search. Yonsook gave a demonstration, showing a wide variety of capabilities. The client is designed to use all the capabilities in the GCMD; in the free text field the user can enter a word for a tag, and then filter down. Agencies can build their own client on top of this using the tag feature.

### ISRO Report

Nitant Dube\* presented a report for the ISRO CWIC partner, giving a status of the integration with CWIC. He began with an overview of the ISRO connector interface with CWIC, and confirmed that the MOSDAC and NRSC data centers are interfaced with CWIC. He listed the metadata and products available in the Meteorological and Oceanographic Satellite data archival Centre (MOSDAC) and National Remote Sensing Centre (NRSC).

Nitant gave a short demonstration of the process of accessing the MOSDAC and NRSC using the CWIC interface, including discovery and data ordering. Future plans include registration of NRSC DIFs (Resourcesat-2 LISS-3, LISS-4 and AWiFS), testing of NRSC Direct Ordering Interface integration into CWIC, and exploring the possibility of enabling OpenSearch for the MOSDAC connector.

Nitant thanked his team and the CWIC team for all their help, and Yonsook observed that it is great to hear their plans.

### ROSCOSMOS Report

Yonsook presented for Tamara Ganina\* on the integration of ROSCOSMOS GeoPortal services and CWIC. She reported improvement of the web search in the IGDIS Integrated Catalog, and its structure and organization of product metadata collections (satellite images). The team is working on mapping of metadata parameters (IGDIS Integrated catalog and CWIC catalog), an an English interface, preparation of data descriptions and their further registration, and on downloading and ordering product. A special link for this purpose will be located in the metadata. Only a limited number of products are available for direct downloading. Plans are underway for release of the new version of the GeoPortal in public access.

Yonsook listed the RSS collections (with dataset and platform) that they will be connecting to CWIC.

### NOAA Report

Ken McDonald\* gave a presentation on CWIC community client development. The motivation is to embark on the new stage of community partners. With the new organizational structure at NOAA, NOAA will begin to link to CWIC. So far NOAA has brokered discovery search and access, developed prototype clients, broadened their offerings and supported CWIC operations.

Ken noted that the CWIC team has been very proactive to encourage community client development; one success is the LSI Portal. The CEOS OpenSearch interface enables common access for community clients to both CWIC and FedEO, so now is a good time for renewed effort. In 2016 the CWIC team will take a more proactive role in developing CWIC clients. The effort will identify partners (VCs, WGs, CEOS Projects, etc.) who are interested in developing clients, explore/analyze past efforts at client development that might provide reusable tools/approaches, and undertake joint efforts with selected partners.

NOAA has identified some resources to support CWIC side of these efforts: Initial interest from the GHRSST Project (preliminary discussions underway), and the team is open to other ideas and suggestions.

Ken commented that this initiative is very similar to the work of GEO. Similar to the CWIC environment, they have the GEOSS portal that provides the access, and they recognize that communities will want to create their own portals. The goal is to maximize utilization of the GCI by promoting and enabling development of community portals/clients and leveraging community infrastructure capabilities. It is an opportunity for WGISS – GEO collaboration.

## OpenSearch Validation

Doug Newman\* gave a presentation of the CEOS OS Validator. The concept is to supply an OSDD location, with a validation tool that performs a series of tests to measure compliance of OpenSearch API. It is categorized by specifications, extensions and best practices, and the results show what is gained from success, and what is lost from failure. The validator checks for OpenSearch Specifications such as Geo extension, time extension, ESIP Best Practices, parameter extension, relevancy extension, OGC Extensions, CEOS Best Practices, and CEOS Developer Guide.

The validator is limited to checking the example query in the OSDD, so the example query should be chosen wisely and as exhaustively as possible. This kind of automated testing is good for first level testing, and second level testing can only be done by a system-level team, to determine if the search results accurate and if the search results conform to “WGISS conventions”.

The validator can be found from the CWIC project page.

## WGISS Connected Data Assets

Yonsook Enloe reported that WGISS has been requested to count the data assets available to CEOS and GEO. Metrics are needed for CEOS Agency data that are searchable and accessible via WGISS supported standards; this is a CEOS contribution to GEOSS. Metrics will be made available for collections (data records of one mission, sensor, and product type and the associated knowledge; an ensemble of some products/granules having a common focus or theme or purpose) and granules (the smallest aggregation of data which is independently managed; a data collection can consists of many granules).

In order to be counted the collection must be registered in the IDN; granule metadata can be searched by independent clients using the WGISS supported standards, provide the data has an access path (no need to be part of CWIC or FedEO). It must also pass first-level system testing to verify conformance to OGC CSW 2.0.2 or the CEOS OpenSearch BP, and second level testing by a system level team to ensure that the search results are accurate and conform to the WGISS conventions.

Yonsook presented the WGISS Connected Data Assets webpage, which will include metrics from IDN, CWIC, FedEO, and independent servers’ assets. A system level team will be necessary for assistance in registering data collections, support of new and existing client and data partners, testing, monitoring, collecting metrics, and planning and implementing the system evolution. This would be a system-level team under WGISS.

Mirko agreed that putting together such a team would be valuable, as this would give a view to the outside world as a whole, a team to help users, and show GEOSS how something can be run well, with a high confidence factor.

Andy observed that WGISS should also work on a BP of CSW, so that agencies follow the conventions that return consistent results. Status codes and exception codes will also help the users. But a specification document is not enough; a system-level team that is constantly working together to tackle the current issues is essential to provide a system view. Andy noted that one of the original ideas was to not have a broad API for everyone but rather a set of standards (CSW, OpenSearch); WGISS provides the system-level perspective. He added that this group can tackle the issue of search relevancy.

Richard observed that in GEODAB, users see that the data comes from CWIC or FedEO, but they should all come from CEOS instead. This needs to be communicated Osamu Ochiai.

**Action WGISS-41-06a: Yonsook Enloe and Mirko Albani to initiate a unified architecture for WGISS Assets.**

**Action WGISS-41-06b: Yonsook Enloe and Mirko Albani to discuss the concept of developing and sustaining a WGISS discovery and access system level team.**

## Access Discussion

The Missions, Instruments, and Measurements (MIM) database was discussed. The MIM lists each mission with a link to the mission page. Yonsook noted that George Dyke asked the agency representatives to revise their information and make corrections. This is an opportunity to improve the users’ access to data by providing the most commonly used link. There is a lot of parameter information contained there that would benefit from having the matching metadata. One of the problems with the MIM is that those parameters are not searchable fields.

In response to a question from Martin it was confirmed that the MIM is considered the official list for CEOS. COVE is connected to it, but they are two separate systems.

**Action WGISS-13-06a: Michelle to send to agency representatives instructions about updating the MIM spreadsheet with URLs to the mission website and a URL to the data access portal for the mission data.**

**Action WGISS-13-06b: Agency representatives to populate the MIM with their changes to the mission website URL and the mission data access portal URL.**

**Action WGISS-13-06c: Andy Mitchell to follow up with the MIM team about suggestions to enhance the MIM and give an update at WGISS-42.**

# Data Use

## COVE Tool

Brian Killough gave a presentation on the CEOS Visualization Environment (COVE) tool, a browser-based suite of tools for searching, analyzing, and visualizing actual and potential satellite sensor coverage. Brian reported that it has a large user base, has been changed from Google Earth to Cesium for global interface, and includes new missions, new overlays, new analysis tools, and new and future archive links. It is hoped that COVEwill become a popular international free/open tool for conducting historic satellite coverage analyses of past and future potential coverage. The tool will include most CEOS missions and connect to every open/free mission archive and links to the data ordering tools of any agency.

The SEO would like to expand the number of links between CEOS missions and the COVE tool. WGISS may be able to help since they have successfully connected to many data archives through CWIC and FedEO. The metadata needed for COVE is scene ID, date and time of acquisition, processing level, scene boundary, cloud cover percentage, solar angles, viewing angles (optical and radar sensors), mode, polarization, ascending/descending.

Brian asked WGISS if there is a unique “ID list” for mission and instrument combinations in CWIC, and if CWIC provides the needed metadata for more missions. Michael and Yonsook will provide the answers to these questions. The next generation of IDN may have a simplified database of metadata.

**Action WGISS-13-06d: Brian to talk with Michael, Yonsook and Mirko about the unique ID for mission and instrument combination in CWIC and FedEO.**

## CEOS Data Cube Prototypes in Kenya and Colombia

Brian Killough gave a presentation on the CEOS DC (Data Cube) prototypes in Kenya and Colombia. The vision is that the CEOS Data Cube infrastructure will become a commonly-used free and open source software toolset for creating local, regional or national pixel-based time-series of multiple datasets that are spatially aligned according to user needs (spatial region, time period, data layers, and grid projection). Users will connect free/open user interface tools to the Data Cube for common analyses (cloud-free mosaics, time series change detection) or utilize APIs to develop their own tools to query the content. Scene-based tools will be used less frequently as users move toward a preference for Data Cubes, and space agencies will systematically supply analysis-ready data (ARD) products that are easily ingested into Data Cubes.

CEOS is working with agencies to develop plans for sustained provision of ARD, with Sentinel-1A and Sentinel-2A as the highest priority. CEOS is also testing prototype Data Cubes for Kenya and Colombia, and testing local, regional hub and cloud deployment. CEOS is developing ingestors to add more datasets, prototype user interfaces for custom mosaic creation, and APIs for users to create their own user interfaces.

Brian described the Kenya and Colombia Pilot projects, which are a custom mosaic tool and a water detection tool. Andy commented that WGISS is keeping a list of open source software, and this would be a perfect example.

Brian discussed ARD, noting that the term has yet to be defined clearly in CEOS, but that most agree on these concepts: ARD is data that has been processed for calibration (e.g. offsets, biases, normalization), corrections (e.g. topography, atmosphere, solar angles, viewing angles) and spatial positioning (e.g. georectification, grid projection) to reduce the computational burden on end users and allow immediate analyses to support user applications. Examples of ARD could be typical “Level-2” products (e.g. optical surface reflectance, radar) or it could be typical “Level-3” products (e.g. mosaics, merged datasets).

Brian suggested the following areas for WGISS support:

* Expand the connections from mission archives to the COVE tool. There are many more archives that could be displayed in COVE.
* Promote the systematic production of “analysis ready” datasets and discuss options for implementation.
* Support the SEO as it investigates BRDF corrections to remove “scene edge effects” from Landsat data. This issue is commonly seen with dark forests and impacts scene classification and is caused by solar angle variation due to varying distance of each pixel from the “center line” location. This may also be an area of support by CEOS WGCV and its LPV subgroup.

Participants commented that there needs to be a scene-based flag at the pixel level.

## Future Architecture for Data

Robert Woodcock began the session with an introduction to the CEOS Chair Initiative. There are three recurring major topics:

* ARD, its role and purpose; taking level 1 data and preparing it for what users want. If this is already done, it becomes more useful.
* Near-real-time – change detection, hazards, monitoring, and impact on cal/val timeliness for ARD.
* Multi-sensor integration, multiple modalities.

Suggested recommendations to CEOS regarding management and utilization of EO data include standardization, interoperability, how CEOS priorities might benefit from the proposed activities, and accounting in regards to use. It would be helpful to identify commonalities that can be provided to the CEOS community; the agencies will make their own decisions, but can be guided by an approach that develops a commonality. There is also a need to recognize where CEOS is today and the best way to get to the next step.

Information can be found at <http://www.datacube.org.au/wiki/Australian_Geoscience_Data_Cube_Wiki>

The document being developed should identify what it take to achieve the recommendations, and discuss handling of near real-time MODIS DC, and the minimum cal/val ancillary data required. It will also discuss ARD, including the case of inland water bodies, land reflectance standardization. This suggests that the report should include aspects of governance.

Robert observed that ARD is very application-specific, and will be very hard to define; a broader definition (minimum set of requirements) is needed. The goal is to remove the burden of processing from the user, providing minimum features. This is an opportunity to communicate recommendations to the space agency leaders. The space agencies need to make provisions to provide ARD. CEOS adopted the processing levels many years ago; what is missing from these? Can there be something in the metadata that identifies ARD. For ARD, Level 2 is the minimum.

For Landsat analysis ready data preparation, the steps include observation acquisition, spatial alignment, observation quality assessment, geophysical parameter derivation, spatial partitioning, and analysis through High Performance Data and Compute. If the space agencies will provide ARD, it will need common formats and definitions, calibration, viewing angle, solar angle, orthorectification, atmospheric correction, pixel-level metadata, and the level of uncertainty of the ARD depending on pre-processing method.

Once ARD is defined, WGISS may be able to tag the collections and files as ARD, and provided the documentation that defines the ARD. ARD will have a significant impact on the discovery system; the metadata will have to be richer than it is.

Robert showed the current state of the document, and it was suggested that timeliness of response, especially for NRT be included. It is also clear that minimum requirements will vary by user community. Where WGISS is mentioned as a contributor for chapter 5, WGISS efforts on big data and cloud computing should be included.

**Action WGISS-13-04:** WGISS to review relevant CEOS and agency activities and ambitions in order to support the study on future data architectures: WGISS efforts on Big Data and Cloud Computing; and Provide Technology Exploration Interest Group summary on past efforts.

## Recovery Observatory Project

Richard Moreno gave a presentation on the Recovery Observatory (RO) project. He described the idea, objectives and rationale of the Recovery Observatory, and showed the timeline of the project and the WGISS involvement.

The main features of the public web portal include editorial content, project presentation, general news and events, collaborative groups to allow users to post information and data, and to provide security management for private and public groups. It also includes maps with product footprints, full resolution products and download, and list of products, quicklook and footprint.

Richard described the technologies and COTS used, and described the product ingestion. The project team is working on adding new features while waiting for a trigger. This project is an example of good collaboration between working groups, and would make a news item for the CEOS webpage.

**Action WGISS-13-09d: WGISS-Exec to provide a news item for the CEOS webpage using the Recovery Observatory as a story of good collaboration between working groups.**

# Technology Exploration

Satoko Miura introduced the Technology Exploration Interest Group session.

## Semantics

Peter Wang gave a presentation on Natural Language Interface for Exploratory Data Analysis, describing the steps that GA, CSIRO, and NCI are taking toward a more natural language interface for exploratory data analysis. He began with a demonstration of the prototype system, performing some exploratory data analysis on Landsat 5 data of Lake Burley Griffin for one year during 1990. The prototype will create tasks to pull band 30, band 40 and pixel quality data, and then create tasks that perform NDVI, adjusting the value, applying cloud masking and performing median reduction summary over time. Peter showed the result of the median reduction over time. The remaining code prepares the data for input into glueviz for further Exploratory Data Analysis. He displayed the slices through time, and displayed a plot of latitude/longitude over time.

Peter raised two questions that are at the very core of analytics:

1. When performing analytics, if you can remove all the things you do not care or want to think about, what do you have left?
2. Should there be a difference between exploratory data analytics and analytics of big data? How much more work should be undertaken to shift between the two?

Three goals for the Australian Geoscience Data Cube (AGDC) have been set:

1. Make analytics simple so anyone can use the DC without asking for assistance.
2. Make analytics for small data and big data appear the same.
3. Make analytics appear the same on desktop, cloud, HPC, GPU.

The team intends to simplify analytics by modularizing and decoupling the analytics from the computation. The Analytics Engine handles the problem description and the Execution Engine handles the distributed computation. The design goal is to enable a separation of concerns and ensure both parties’ concerns are addressable. The goal is to separate the analytics from “dishing out the computation” so EO scientists do not need to know about the distribution of the computation. To make this possible a good interface is needed; these are the Analytics API and the Analytics Expression Language.

The Analytics API consists of 3 components:

1. Analytics Expression Language is a customizable math-like language that can be used to express the analytics.
2. Analytics Engine is responsible for encapsulating the data and task description and its dependencies into a workflow.
3. Execution Engine is responsible for platform independent execution of the workflow.

The Analytics Expression Language is a mini intuitive language to express the analytics; an expressive n-dimensional array abstraction interface with which scientists are already familiar; it is something that looks like a matrix or a multi-dimensional array. From a software perspective it is an abstraction and a way to describe the data and the analytics. It has information about the data and what it looks like, how and where to get the data, and the transformations that will happens to the data.

The Analytics API has 2 main functions.

1. create\_array – this creates a virtual array or if you’d like a description of the data, what it looks like, how to retrieve it and other useful metadata (i.e. enough data to be able to know if x is applied to it, what it will look like).
2. apply\_expression – this is the task creation step that takes as input the output of create\_array and some analytics expression and creates an analytics task.

How can it be made more natural language-like?

1. Explore existing W3C standards for encoding the workflow.
2. Explore linked data as part of the semantic web.
3. Explore Natural Language Processing techniques for parsing and making sense of plain English queries.

Peter observed that the goal is to develop methods and a reference implementation for a linked data DC. It lines up beautifully with the development of the Australian Geoscience Data Cube as a storage and access model (this is not by chance). Further exploration may occur as a way to increase interoperability of the AGDC.

In response to a question from Satoko, Robert replied that the API is the second version of the data cube. It will have seamless action with no view of tiles or files and seamless on the storage boundaries. Richard observed that in the end you want to use natural language to make an analysis.

## Big Data

### Bridging the Gap between HPC and High Performance Data Analysis

Lesley Wyborn gave a presentation on Bridging the Gap between HPC and High Performance Data Analysis. She began with a description of NCI HPC and HPD integrated activities, and a listing of their high performance data collections. NCI’s goal is to enable global- and continental-scale as well as scale-down to local/catchment/plot scale of NWP, climate, ocean, earth system, water, vegetation. Some of the foci for satellite imagery include modelling extreme and high impact events, NWP, climate coupled systems, hazards, and monitoring of the environment and ocean. She showed a graph of I/O speeds per packet for UM files with and without MPI-IO. She also described NCI’s National Environmental Data Interoperability Research Platform (NERDIP).

Workflow engines, virtual laboratories, and science gateways are infrastructures to lower barriers to entry. They need to support grid diversity since downstream communities may not wish to deal with different grids, but the modelling communities generate data appropriate to them.

Lesley described data file layouts and performance analysis, where data is stored in chunks of predetermined size; a two-dimensional instance may be referred to as data tiling, and chunks are matched to the cache size of the processor. She described benchmark testing of a Landsat scene in various file formats, finding that accessing chunked/tiled datasets is faster than accessing contiguous datasets. She graphically displayed results of read performance vs. file size. And observed that another finding is that Dynamic Web Map Tile Servers with THREDDS Server is effective. Lesley gave an example of re-projecting raster data on-the-fly from multiple satellites.

### Recent Work on Remote Sensing Cloud Computing and Big Data Processing

Lizhe Wang gave a presentation of recent work on remote sensing cloud computing and big data processing. He began with a discussion of the data volumes involved, highlighting the volumes and data rates of data acquisition, and throughout the entire flow. He also discussed the complexity of the processing and extraction algorithms.

Lizhe described the software systems for high performance RS big data processing, which include:

* Global Distributed RS Data Model in Memory
* Generic Parallel Processing Model for RS Big Data
* Parallel Storage Model for RS Big Data
* Scheduling of Large-scale RS images Parallel Processing

Current work involves a multiple data center collaborative process system. It includes 40+ types of RS inversion products: vegetation structure and growth state, radiation budget, water and heat fluxes, ice-snow variation, mineral exploration, and 20+ types of thematic RS products: agriculture, forestry, geology, water, sea ice, environment. The system scale can be global, China, South America, and use annual time series. The service provides spatial metadata management for multi-level RS data, data-dependent knowledge base, replicas management for distributed RS data, automated data production, production task scheduling, workflow fault-tolerance, and distributed resource monitoring. Lizhe displayed diagrams of the software system architecture, and of file, metadata, workflow, and replica and cache management.

The findings are that big data is computationally and data intensive, with unbalanced data flow and complex algorithm flow. He recommends specially optimized HPC software systems for data transfer, IO, storage, processing and management.

Satoko asked where the data processing occurs; Lizhe replied that it is near data center.

### NoSQL Influenced Solution for EO Data Exploration

Jeremy Hooke gave a presentation on a NoSQL influenced solution for EO data exploration. He began saying that in version 1.0 of the AGDC they used the traditional relational approach, where there is a table for every entity and a column for every field. This approach quickly becomes complex and is not very extensible. Postgres introduced composite types several versions ago. One column can contain many values, and more recently, composite type JSONB. Supporting N-dimensions traditionally with two tables is also very complex.

N-dimensions with composite types support one table; this means no joins, readable queries, data locality, and fields can be indexed and constrained. Going a step further, JSON-compatible metadata documents can be put directly in the database as a JSONB column, and then index and search the fields directly from it. This simplifies the schema immensely, but it is entirely up to the code to maintain the field structure, and it is limited to JSON datatypes. The code also needs to manage foreign keys.

In a hybrid approach, concepts are defined by users (dataset metadata, sensor information, storage configuration) and by the core developers (concepts of dataset, storage unit, ids, relationships.)

Andy asked if ingestion of new data into the data cube is already happening; Jeremy replied that they have built a proof-of-concept for this; the metadata model is not fixed. Lesley added that there will be a minimum metadata standard that the group follows.

Robert summarized that:

* ​​AGDC metadata is for the operation of the AGDC analytical capabilities in a closed system and should not be confused with ISO and other metadata standards whose purpose is for data discovery across multiple organizations - one is internal, the other is external (open)
* AGDC metadata would map to the ISO and other discovery standards  - it would more than likely be a subset and lossy mapping from AGDC to the discovery standards, in particular some aspects would become natural language descriptions in the discovery standards because of their inability to deal with fine granularity of sensor specific differences
* The NoSQL approach being used in AGDC allows for easier implementation of the finer differences between data types (e.g. SAR vs optical satellites). Aspects of the metadata are common but you can be very specific  with differences that matter,  which is particularly important in analysis (which is the next level of use following discovery)

### Discrete Global Grid Systems: A New Way to Manage 'Big Earth Data’

Matthew Purss gave a presentation on Discrete Global Grid Systems (DGGS) as a new way to manage 'Big Earth Data’. He began with the challenges of big data, and described a DGGS. A DGGS is a Digital Earth reference model designed to be an information grid, not a navigation Grid. OGC defines a DGGS as:

“…a spatial reference system that uses a hierarchical tessellation of cells to partition and address the globe. DGGS are characterized by the properties of their cell structure, geo-encoding, quantization strategy and associated mathematical functions.”

Any tessellation of the Earth does not necessarily produce a DGGS, and single resolution computational grids are not sufficient to constitute a DGGS. A DDGS must have a hierarchy of grids with successively finer resolution. Global Grids that do not have ‘Equal Area’ cells are not sufficient to be described as a DGGS. Equal Area cells are critical to ensure uniform coverage of data in an area of interest, to facilitate efficient statistical analyses, and enable standardized interoperability between DGGS and other data infrastructures.

Equal Area Partitioning is achieved with Platonic Solids mapped to the Surface Model of the Earth. Unique Cell Indices are hierarchy-based, space-filling curve, and axes-based or encoded address. Matthew displayed a taxonomy tree showing the different styles of global Grid.

The diversity, incongruity and lack of standardized applications of global grid infrastructures limits the development of accurate analysis tools for Big Earth Data, so standardizing is important. The OGC established a Standards Working Group to address this problem; standards adoption is anticipated in mid-2016.

A DGGS provides a framework where the three fundamental questions of geospatial analysis can be answered:

* “Where is it?”
* “What is here?”, and,
* “How has it changed?”

Big Earth Data that is aligned to a DGGS is easy to access, store, sort, process, transmit, integrate, visualize, analyse and model.

Arnold Dekker asked how they deal with mountains and deep troughs. Matthew replied that DGGS is built on the Geod model, so it dealt with part of the complexity in the background. The next step is codifying the easier form of DGGS, with later iterations to factor in other dimensions (temporal, height).

Ben asked if they handle scalar or vector fields. Matthew replied that it is immaterial whether it is scalar, vector, or raster, if you have accuracy of the position. The observation is stored as an observation in space and time, and integrated analysis can be done without having to convert. The DGGS provides the common framework.

Robert asked if they have looked at the requirements of model data simulation; this problem is exactly why there are multiple styles of grids. Matthew noted that you choose the DGGS that fits the geometry. Through the index mapping you can extract the results without having to convert; commonalities open the door to analysis.

Tom Cecere asked about handling Earth movement due to plate tectonics; Matthew conceded that this is a fundamental dilemma; most conventional applications have been around fixing a coordinated space that is tessellated. With DGGS the data will shift through the data store, as long as there is good control of the spatial and temporal observations. A DGGS on a fixed coordinate framework allows this.

### Storage Structure and Efficient File Access

Joshua Sixsmith gave a presentation on storage structure and efficient file access. He began discussing aspects and measures of efficiency and described physical aspects of the storage medium. The region of interest selected should test a variety of conditions; sparse, dense, as well as a variety of land-covers, which all will have some form of impact on compression.

Joshua displayed results of read and compression efficiency averaged across each cell. The LZF compression resulted in faster read times than the raw data. Smaller chunk sizes resulted in slower read times. Chunks are not a factor of the array dimensions (larger files, slightly longer read times).

Joshua discussed a simulation with synthetic data, and appending newly acquired data (collection management). He showed the results of dataset creation times, comparing simple append, and append in chunks. Dataset create times, GZip and LZF vs append in chunks, and compression times went down dramatically. He also displayed synthetic algorithm process times comparison, which showed bad performance. Care should be taken with the algorithm selected.

Joshua concluded with the following key points:

* Write in chunks that contain all data required; this aids in collection management in that a file could be aggregated when there is sufficient data required by the chunk dimensions
* The main type of analysis should determine how the data is chunked; and an in-between solution can be chosen.
* Large chunks can improve read times (but can also be a downside if you require less data than that returned by a chunk).

### Bringing the Processing Close to the Data

Richard Moreno gave a presentation of bringing the processing close to the data, within the context of big data: big volume and big processing. Limitations are cost of storage of several PB, and bandwidth resources, but there is no more downloading of the full archive or bulk extraction: the processing occurs close to the data within the context of high data volumes and processing, and changes in data usage. The future may hold integration of Copernicus mirrors and the possibility to federate DCs.

Significant advances in data downloading have occurred due to new tools, standards, interoperability, web services access (OpenSearch), and bulk extraction (Metalink, JDownloader). But these do not solve the bandwidth resource problem, and the duplication of storage.

Processing close to the data can occur with interactive processing via web services (WPS), interactive processing via MMI (Google engine, GA Analytics Expression language, Notebook), mass processing on HPC/Cloud, and SandBox for algorithms/processing tuning.

Richard concluded that this talk is meant to stimulate these discussion points: Is big data compatible with distributed architecture, for example OGC OWS-10, ESA and European agencies Federated pilot, and EUCLID project? Can this be generalized? Is a centralized platform / cloud the unique solution? What are the advantages and disadvantages of cloud computing based architecture?

## Technology Exploration Discussion: Semantics and Big Data

Satoko discussed a proposal for Technology Exploration webinars, a series of monthly webinars, each 30-60 minutes, on various topics. Each webinar will be conducted by an expert in the specific topic, and include a wiki page for each month advertising the expert speaker, a description of the talk and the logistics of the webinar. The webinars will be recorded, and the recordings will be available on the WGISS website. The speakers will have expertise in the subject and will come from a variety of international organizations (no CEOS affiliation needed). The following is a preliminary list of topics:

* Big Data, HPC, and Cloud Computing
* Searching for free satellite data from CEOS agencies
* GCMD/IDN Keywords – what are they, how to add to these lists
* Search Relevancy for collection searches
* Data Quality (IOS?)
* Semantics, augmentation of metadata
* Visualization of Earth Data (web-based, with volume rendering, tiling, and augmented reality)
* User authentication, single sign-on
* Metrics of usage, datasets
* Crowd sourcing

Tom suggested that some of these topics could be focused to the specific needs of the CEOS community, and Andy noted the issues with distributed architecture with processing occurring at distributed data centers. Tom added the challenge of not just finding the data, but getting it in a suitable form and how WGISS can help defining the ingest format. Lesley suggested identifying the locations and availability of processing centers, and CEOS-related big data issues. She added that there are organizations (RDA) where the discussion is going on. Ben suggested adding APIs and use of standards under Big Data and Cloud Computing. Andy added the topic of networks, and how bandwidth affects certain users. Richard suggested the advantages and disadvantages of cloud computing.

Under visualization of data it was clarified that it includes time series visualization, what methods will help; Nitant added volume rendering, tiling and caching, visualization technics, and web-based visualization. Andy raised the issue of authentication layers, federated single-sign-on, federated metrics collection, and augmented reality. Jono suggested adding harmonization and volunteered to be a contributor. Richard mentioned data cubes, and suggested definition of open data.

Tom suggested relating the webinars to the WGISS work plan to assist prioritization. Richard said the maturity of the topic comes into play in deciding where to start. Richard suggested that the presentation he made at the side meeting two years ago would be a good start.

**Action WGISS-41-07:** WGISS to develop monthly webinars; several topics have been suggested and should be prioritized.

# WGISS/WGCV Joint Plenary

## Welcome and Introductions

Andy and Albrecht von Bargen welcomed the WGCV and WGISS participants to this joint session, and thanked CSIRO and GA for hosting, and Alla Metlenko and Medhavy Thankappan for all their hard work facilitating the meeting.

## GA Welcome Address

Dr. Chris Pigram (CEO of GA) welcomed the participants on behalf of Geoscience Australia. He highlighted the importance of satellite observations to Australia; though it has a very low population density, it has the third highest marine jurisdiction. The data from Earth Observation supports over 100 government programs, and Australia is entirely dependent on data from the global community and as a result desires to reciprocate in a beneficial way. This is done through bilateral engagements and through organizations like GEO and CEOS and assisting with ground-based infrastructure.

Australia has a long-standing partnership with USGS; since 1979 they have operated a ground station to receive Landsat data; their new Alice Springs station will take this support to a new level, with a world class system. Australia also has a partnership with ESA for regional analysis hub for the SE Asia region.

Australia also has a long history of supporting calibration and validation. The infrastructure in Queensland has provided stable calibration targets in many campaigns, and they look forward to supporting further capabilities.

Australia is committed to make satellite data easy to use at large scales, comparable and analysis ready, and to support time series analysis, and to provide the best available science. The Data Cube version 1 showed that this was possible, and is already being used for water management. Version 2, due for release soon, will provide products for decision makers. In line with supporting the global community, they are releasing it with an open license.

A key point is Australia’s commitment to open science, releasing everything with open licenses. Involvement with CEOS allows work with a common set of goals. Subgroups like WGISS and WGCV ensure that the data actually gets used, providing a seamless experience to users, and a roadmap to the future.

Dr. Pigram invited the participants to tour the GA facilities, and look forward to hosting a cocktail party in the evening. It is his pleasure to open this joint session.

## CSIRO Welcome Address

Dr. Alex Held welcomed the participants on behalf of Dr. David Williams and CSIRO. CSIRO is happy to help bring their teams together with the two CEOS working groups. He began with a briefing of CSIRO.

CSIRO is Australia’s science agency, an innovation catalyst. He described the scope and organization of CSIRO, highlighting the business unit and application focus areas. CSIRO’s business and focus areas include all the obvious areas, but also all includes national facilities, for astronomy and space science, animal health laboratories, national research collections, marine national facility, and National Computing Infrastructure (NCI). CSIRO’s work in radio astronomy has global significance, due to the isolation of the infrastructure.

Alex listed the areas of international participation, with specific mission science teams, global initiatives like GEOGLAM. He noted a recent report on Australia’s cal/val capability, with laboratories in East and West Australia having capabilities for long term cal/val. They conduct a number of land and water validation studies, including testing robotic infrastructure, capitalizing on the varied vegetation architecture.

Alex gave background on CEOS, including membership, organization, and objectives. He describe the Strategic Implementation Team (SIT), the Virtual Constellations (VCs), the working groups, and the current work plan. He described the expected outcomes in terms of Climate Monitoring, Research, and Services, Carbon Observations, Including Forested Regions, Observations for Agriculture, Disasters and Water, Capacity Building, Data Access, Availability and Quality, Advancement of the CEOS Virtual Constellations, Support to Other Key Stakeholder Initiatives, Outreach to Key Stakeholders, and organizational issues.

He listed the current and planned CEOS agency missions, and CEOS participation in GEO.

## CEOS Executive Officer Report

Jonathon (Jono) Ross gave the CEOS Executive Officer (CEO) report. The role of the CEO is to support the CEOS Chair and the SIT Chair and to coordinate with the CEOS working groups.

He gave highlights of the 29th CEOS Plenary, noting that JAXA did a brilliant job of coordination and hosting in Kyoto. The Kyoto statement summarized the meeting:

Applications: significant progress but still work to do.

Engagement with GCOS and outcomes from Paris climate conference. New GCOS Implementation Plan, with emphasis that this is something worth doing.

Sendai Framework: GEO-DARMA

Sustainable development: connecting with the statistics community, to make the data useful.

GEO: engage to make the next decade a success. Build on key successes like GFOI, GEOGLAM, AfriGEOSS, Blue Planet, Carbon strategy, and Water strategy.

New members of CEOS were welcomed: Malaysia, Gabon, and Mexico. The CEOS Strategy for Water Observations from Space was endorsed, and the Water Constellation Feasibility Study is now in train. The disasters pilot activity was extended to cover landslides, and to further support the super-sites. CEOS confirmed continued support to GFOI and GEOGLAM, and revitalized the LSI-VC, and the efforts toward Analysis-Ready Data (ARD) to capitalize on massive free data streams. Many are calling for these data assets, and a lot of agencies are stepping agencies to support this.

The current Chair Initiatives center on future data access and analysis architectures, and non-meteorological applications for next generation geostationary satellites. The “Big Data” challenge is a new paradigm, and an opportunity for enhanced integrated modeling, archiving, and platform technologies to enable use and access. He highlighted the importance of this based on the range of internet access speeds around the world, and on the estimated growth of data storage needs.

The resulting study of data access and analysis will highlight what agencies are doing today, what lessons can be learned from the prototypes currently underway with the governments of Kenya and Colombia. It will also confirm issues and opportunities resulting from the trend towards Big Data and ARD, and what issues should be worked on cooperatively. Jono discussed non-meteorological applications, giving examples, and listing opportunities for member agencies.

The 2015-2017 CEOS Work Plan outcomes are reasonably on track. He presented the timeline for the 2016-2018 Work Plan, whose release has been delayed by about one month. Highlights from this work plan included that water actions are shifting from study to implementation, following endorsement of the CEOS Strategy for Water Observations from Space. The CEOS Chair initiatives are incorporated into NG-1 and DATA-7. The climate (CMRS) shifts into second cycle of ECV Inventory and will focus on new GCOS-IP, Satellite Supplement and Response. GEO-DARMA is recognized in DIS-15. WGCapD activities are significantly recast to reflect new strategic objectives, and the CEOS Data Cube is explicitly included under the leadership of the SEO.

GEO-XII confirmed strong support for the next decade of GEO, with 400 attendees. The GEO Strategic Plan 2016-2025 was endorsed by the GEO Ministers. New rules of procedure were adopted, addressing CEOS concerns around Programme Board and ExCom representation, as was the GEO -2016 transitional work programme. GD-5 is led by CEOS, and “Space Tasks” are included. The CEOS booth was well attended with significant interest around future data architectures. Jono thanked the team who worked on that.

Jono noted that GEO has named a Programme Board, with the role of developing and monitoring delivery of the 3-year GEO Work Programme, and has also added three participating organizations (including CEOS) to serve as Observers on the ExCom. GEO now has a big focus on delivering impact into big global and regional agendas, in particular the Sustainable Development Goals. Focus on developing stronger and more explicit links with UN institutions and programmes, and on mobilizing resources from private sector, development banks, and philanthropic organizations. There will be an emphasis on defining and prioritizing requirements by SBA, with a strong focus on in-situ coordination and space/in-situ integration. At the annual CEOS-GEO Coordination Meeting, CEOS-GEO will discuss collaboration around WIGOS to explore ways to strengthen their partnership.

The SIT Chair priorities are to ensure the success of ongoing CEOS priorities, to ensure full access to and exploitation of Copernicus Sentinel data, and to further develop the CEOS-GCOS-IPCC-UNFCCC relationship for climate measures. The CEOS global goals for sustainable development were adopted in 2015. The role for EO is identified in tracking and achieving progress, working with UN-GGIM and GEO to promote the role of EO in indicators framework.

CEOS Experts are intended to ensure that CEOS has visibility in key domains where a WG/VC exists. The role is formalized through invitations from CEOS chair, and responsible for watching brief, and advising on emerging issues, input to the work plan, and status tracking.

Jono thanked WGISS for submitting its Terms of Reference document on time.

Andy inquired about more coordinated effort with other WGs. Jono suggested looking at the WGClimate work plan to get insights as to where WGISS can participate, and Jono offered to facilitate the communication. Stuart noted that the working groups really are not connected except where a given agency is supporting more than one WG.

Andy raised the concern that WG reporting is minimal at SIT meetings, so the CEOS principals are not as aware of what the WGs are doing. Alex suggested that the second SIT meeting of the year (the September meeting) is more appropriate for WG reporting, and Jono suggested that a better approach is for a WG to address a specific major topic. Albrecht added that it is difficult to report at the CEOS plenary and agreed that reporting on special topics may be more positive, and the SIT Technical Workshop is a good place to begin. Senior delegates need to be encouraged to attend the WG side meetings.

A WGCV participant noted that some communities do not see the connection of satellite efforts to solve some of their larger problems. Stuart clarified that sustainable development goals are important, and barriers need to be lowered to take the measurements and transform these into useful products. The challenge in the engagement is that the UN has had to act quickly to establish their targets, and these are inadequate. The existing approach is and always will be inadequate. EO has a role to play, and cooperative work should continue in order to progress to the point of something that can work. The statistical community can then get involved, and needs to be engaged. A two-pronged approach where developed countries can provide the products, and developing countries can rely on them is the ideal scenario.

## GEO Secretariat Report

Mirko Albani gave the GEO Secretariat report on behalf of Osamu Ochiai. He gave an introduction to GEO, and presented the EO and Geospatial Information (GI) support to the Sustainable Development Goals (SDGs). He displayed GCI Providers evolution and statistics, showing increases due to the introduction of the brokering approach for discoverability and access, and in the number of brokered enterprise systems.

He explained the Work Programme, noting four types of GEO activities: Community Activities, Initiatives, Flagships, and Foundation Tasks. There are 11 foundational tasks, some very much linked to CEOS. He highlighted GD-02 and GD-07 (GCI Development and Operation), and added that GD-05 and GD-08 are also important to CEOS, and WGCV is working with Ivan Petiteville to provide the cal/val information.

Stuart noted that the conversation and interaction build the understanding of the uses of the data and for the providers to develop it, and facilitates a deliberate conversation with the user community. Greg Stensaas added that there is a lot going on in terms of assessment processes for GD-08, and there may be some benefit in EO assessment.

Mirko observed that CEOS provided more than half of the 200 million GEOSS resources through CWIC and FedEO interoperability agreements. There are more space agencies who have not yet linked to CWIC or FedEO, and CEOS is working to encourage those space agencies to connect to GCI through one of them. It is also helpful that CEOS is joining the GCI development activities (GD07) such as GEOSS architecture evolution on what kind of new services will be foreseen (e.g., access and processing services, big data, data integration) and also join to test the new functionality of the information technologies. GD-07 has six major sub-tasks. CEOS contribution in several of them.

**Action WGISS-41-10a:** GEO Foundational Task GD-03 to identify where WGISS can contribute.

**Action WGISS-41-10a:** GEO Foundational Task GD-08 to identify where WGISS can contribute, and propose this to the CEO, as it is linked with G05.

## GEO Work Plan IN-02-C1 / GEO Strategic Plan GD-05

Mirko introduced the GEO Work Plan saying that before 2015, IN-01 dealt with EO systems; this is now mostly reflected in GD-05. IN-02 advances in data lifecycle management. The priority actions are meant to improve data management, improve quality, strategy for long-term preservation, ensure availability and build on DIAS. IN-03 dealt with the GCI, which is now split into two tasks, one for development and one for operational. IN-05 dealt with GEOSS design and interoperability.

The new GD-07 has different components, and many of them are the old IN tasks, including IN-03.The data management part of IN-02-C1 is now in GD-07.03. What is not evident in is the strategy for quality and the support for the cal/val activities. It also does not make sense to have DMP in a task on GCI development, which is about data delivery. In conclusion, the data management principles and the quality are not completely covered in the foundational tasks.

WGISS has identified its contribution, and provided the approach to the GEO-SEC, but needs add contributions to GD-08 and GD-03, and to ensure that the CEOS activities are properly reflected in the work plan. Jono commented that WGISS will attend the WIGOS side-meeting in Geneva, and that CEOS looks to WGCV and WGISS to provide the data quality and data access.

Kurt Thorne (WGCV) agreed to go through the foundational tasks to determine where they can contribute, and WGISS will do the same specifically with GD-03. In response to a question from Andy, Brian said that GD-05 is more about providing the data, and ends there; GD-07 is more about serving the data; the outcomes of GD-07 will then apply to GD-03.

The Data Management Principles are very high level, and understandable to non-technical readers. GD-07.03 will re-draft these, drilling down to more technical detail. There are a couple of principles that apply to quality. The group recommended to Osamu is to move GD-07.03 out of GD-07, making it its own foundational task. WGCV will comment on GD-07.03 in the data quality principles.

For CEOS representation in GD-08, CEOS needs to understand the task, and then submit feedback to the CEO on potential participation, noting that it is very much linked with GD-05.

It was also suggested to merge GD07.05 into 01.

**Action WGISS-41-11a:** WGISS Exec to suggest to CEO to merge GEO Foundational Subtask GD07.05 into GD07.01.

**Action WGISS-41-11b:** WGISS Exec to suggest to CEO to remove GEO Foundational Subtask GD07.03 from GEO Foundational Task GD07, and make it a separate Foundational Task or move into GD01.

## Introduction on WGCV/WGISS Joint Efforts for Future Data Access

Thomas Schroeder gave a presentation on non-meteorological applications (NMA) for next generation geostationary satellites. He began stating that, as Chair of CEOS, CSIRO proposed to provide leadership and coordination on two new initiatives:

A study of Future Data Access and Analysis Architectures (co-chaired CSIRO and USGS)

A study of non-meteorological applications for next generation geostationary satellites

The task team for this study was established in January for a one year activity period and is supported by 15 CEOS agencies. The study will result in a report that provides comprehensive and pragmatic guidance to CEOS on new opportunities arising from next generation geostationary satellites and GEO-LEO synergies.

Thomas explained that the Himawari satellite provides full coverage of Australia every 10 minutes, so they collaborated with Japan to exchange information. An outcome of the study is a catalogue of non-meteorological products, which he displayed in a table. The non-met applications of Himawari-8 are abundant, in the three categories of land, ocean and atmosphere. For example tidal effects can be studied and visualized with more spectral bands and more frequent observations.

Envisioned NMA study outcomes are:

1. Trends and Outlook for Geostationary EO Satellite Capabilities – Catalogue of CEOS/CGMS agency missions, instruments, measurements, data volumes etc.
2. Inventory of relevant non-met applications and review of initiatives being undertaken by CEOS and related agencies – atmosphere, land and ocean (build on Japan-Australia bilateral effort initiated in 2015)
3. Benefits of synergistic use of GEO-LEO systems
4. Identify key opportunities and benefits to CEOS and CGMS agencies as well as challenges
5. Recommendations for the way forward tactically and strategically for CEOS and its agencies

Challenges and opportunities for CEOS result from expansion and diversification of constellations of EO satellites, and for the need for inter-calibration, data harmonization, inter-operability of data sets.

The NMA co-chairs are Ian Grant (ABoM), Kenneth Holmlund (EUMETSAT), Satya Kalluri (NOAA), and Thomas Schroeder (CSIRO).

Andy asked how WGISS could provide input on uniform data access. Thomas replied that they are not likely to be addressing standards, especially with operational systems; this is just a pre-study to get a CEOS picture of what is possible. The team will definitely seek WGISS input on the next iteration. The future of land remote sensing is hourly 1km observations. The study will point to potentials and challenges. NOAA is going toward enterprise algorithms, and needs to been seen as an example of where there future needs are, and what toolboxes can be used for making the data more accessible.

Tom Cecere co-chair with Robert Woodcock of the Future Data Architectures (FDA) and analysis ad-hoc team noted that the timing of this meeting is perfect because it helps to crystallize their strategy, adding that the study most likely will last more than a one year. CEOS agencies want their data to be accessed by as many users as possible, thus “Enabling Analytics for the Masses”. CEOS agencies want to understand the user community and review the user metrics in order to determine what needs to continue and what needs to change, and how to serve the communities equally. The concept of free and open data changed the paradigm for data access.

Serving the water community may be a good place to start, since it may be advisable to start simple and learn from the simple applications. ARD has many possible variations. The topic is heavy on cal/val needs; who is going to determine and characterize the uncertainties, and provide uniformity.

Tom explained the structure to the report. It will contain the trends, lessons learned, next steps and conclusions, and may address discrete gridding systems.

Albrecht commented that the subject matter is very broad, making it difficult to identify WGCV involvement. He suggested a more structured setup for working together.

Robert replied that there is pressure because of the data volume and the sheer number of applications. They will have to identify the top levels, and allow the details to develop. Currently the thoughts are crystallizing, but in the next months they will start filling in the gaps. He commented that it is NRT applications where the calibration information is not available quickly enough, putting pressure on the data stream; he wondered if that is pressure point for the cal/val community. Albrecht replied that it depends on the communities. It is a matter of disclosure: the calibration can be predicted, provided that the associated uncertainty is presented. To get the best data scientists have to go to the climate data. Alex Held commented that the core datasets need to be identified to determine the intersection between the two working groups.

## Carbon Action Items

Albrecht von Bargen displayed the list of CEOS Carbon Actions, highlighting those that affect WGISS and WGCV. These actions originate the carbon strategy implementation study team formed in 2014, based on the strategy approved by GEO in 2014.

Kurt Thorne noted that 14 of the CEOS Carbon actions are related to WGCV, and there is overlap among many of them. Treating the actions as a discrete set is not efficient due to limited resources for WGCV members, and there is significant overlap in some of the actions. It was necessary to develop an efficient way to close the Carbon Actions, identifying the overlap allows progress on multiple Carbon Actions through completion of a single WGCV action. One approach was to combine the in situ WGCV carbon actions. In some cases WGCV needs advice on actions, or needs information. There also needs to be an effort to inform CEOS of the work that is done. Kurt showed examples of how Carbon 7 and Carbon 34 can be detailed and assigned.

Andy noted that WGISS is trying to identify a time-frame for the effort, and wondered when WGCV would be contacting WGISS on suggested involvement. Kurt said it will be after the next WGCV Plenary (September 2016). Andy suggested that this is a good opportunity to learn what each other is doing. Andy pointed out that WGISS is assigned to a few of the carbon actions (22, 3, 30, 35), and others do not mention WGISS but WGISS obviously needs to be involved. Kurt replied that the task lead is responsible for contacting the WG to help. Albrecht agreed that the SIT team expects the lead to contact the contributing entities.

Tom Cecere raised the subject of the LSI VC. He noted the number of carbon tasks that LSI leads, and pointed out that they would be reaching out to WGISS and others where they can contribute. He added that the LSI will work at a slightly more strategic level, and is also looking at the definition of ARD and to work collaboratively on data cubes. They are also trying to understand 08-02.

Carbon-08-35 is assigned to WGISS as the lead. Stephen Briggs suggested that WGISS host a carbon workshop to fulfill the action. Kurt will communicate to Andy how WGCV best suggests to approach this action. This action is a major outreach between two communities and must be broken into manageable segments.

WGISS agreed to present jointly with WGCV at the 31st CEOS SIT Meeting (April 19-20) how the two WGs are performing an analysis of the Carbon Action items in order to sufficiently address them.

**Action WGISS-41-01:** WGISS Exec to present jointly with WGCV at the 31st CEOS SIT Meeting (April 19-20) how the two WGs are performing an analysis of the Carbon Action items in order to sufficiently address them.

## Discussion on Joint Interaction

Andy presented a list of proposed WGCV/WGISS joint activities:

* Support data access of CEOS Test Site information (CEOS IVOS sites, LPV sites, SAR, DEM and others). WGCV can provide necessary information to WGISS (IDN/CWIC/FedEO). WGCV has previously provided WGISS with several records (< 50) to be cataloged in the IDN. CWIC provides access to several of these records. Continue to update these records
* Support Metadata Requirements for Quality (Quality fields): Review and comment on GEOSS Data Quality Requirements (DDQ -Documentation on Data Quality). Include validation metrics (possibly calibration metrics and sensor performance indicator) Make the reference to validation guidelines into the metadata information. Develop a White Paper on Recommendations for improving Data Quality in Metadata.
* Develop Best Practices on Data Reprocessing – including current and future practices (i.e. on-demand ARD Processing)

The metadata activity would involve validation metrics: how they are defined, and what is the terminology at the various stages of processing. The goal would be to have a good representation to get a generalized approach. Part of the role of WGISS is to make recommendations to the ISO board; but WGISS needs input from WGCV to develop the metadata requirements. The hope is that the user can go to the metadata to make decisions about what to use. Another issue for the user is how know that reprocessing is occurring, since this is not in the metadata.

**Action WGISS-41-02: In order to support, with WGCV, Metadata Requirements for Quality (Quality fields), Andy Mitchell to confirm with Albrecht if these are activities WGISS should pursue:**

**Action WGISS-41-02a: Review and comment on GEOSS Data Quality Requirements (DDQ -Documentation on Data Quality)**

**Action WGISS-41-02b: Include validation metrics (possibly calibration metrics and sensor performance indicator). Include the reference to validation guidelines into the metadata information.**

**Action WGISS-41-02c: Develop a White Paper on Recommendations for improving Data Quality in Metadata.**

**Action WGISS-41-03: Andy Mitchell to confirm with Albrecht the possibility to work with WGCV to develop Best Practices on Data Reprocessing – including current and future practices (i.e. on-demand ARD Processing).**

# Agency and Liaison Reports

## Geoscience Australia (GA) and Commonwealth Scientific and Industrial Research Organisation (CSIRO)

Simon Oliver presented a report on GA and CSIRO. He began with a discussion of Australia’s Regional Copernicus Data Access/Analysis Hub. This is an integrated approach to support the broader objective of enhancing access to satellite EO data by research, industry and civil society, and to facilitate collaboration between Australians, Europeans and inhabitants of the South-East Asia-South Pacific region in the exploitation of Earth observation data. He diagrammed the process leading toward operational products, and for delivery to users and storage in Deep National Archives. One of the operational products is to do with hazards for immediate use.

The UK Catapult collaborative agreement was signed by GA, UNSW, CSIRO, and UK Catapult for the RADAR cube. The data comes from Sentinel, ERS-1/2 SAR, ENVISAT ASAR, and ALOS-1 PALSAR and is a pilot project for Australia region. It has potential for helping with emerging industry. Mirko asked if WGISS can support the pilot; Simon replied that there are a couple of potential application areas. Sentinel-1 data was collected specifically for this project.

Simon gave an overview of the AGDC programme. He began with a description of Landsat ARD preparation, which includes observation acquisition of raw data, spatial alignment, spectral alignment, observation quality assessment, and biophysical parameter derivation. One application is using models to map tidal extents using time series analysis. Another is a study of mangrove change to study the impacts of environmental change. Other terrestrial applications are vicarious calibration sites, Landsat MODIS blending, GEOGLAM Rangeland and Pasture Productivity, and solutions to assist algal management in lakes and rivers; Sentinel-2 is of interest.

Simon described the processing pathway and decision support tools, and version 2 of the AG Data Cube, which is a complete rewrite of v1, to be deployed in July. It includes reprocessing of Landsat, refinements of MODIS, and AVHRR. A key aspect is the handling of different grids and formats, letting the user control the process, processing to certain indices separately, and then blending to the indices.

Simon also described the AGDC v2 data ingestion, which is open source. DGGS an important part of the future versions of DataCube.

Future WGISS involvement may be:

* AWS - S3 object storage study: want to move to the Amazon environment. SEO also has a proposal with Amazon. Andy said it would be nice to bring the results of everyone’s studies. A lot of people are looking into multi-dimensional storage in the cloud environment.
* The Future: Discrete Global Grid Systems - OGC SWG and implications for AGDC evolution.

Richard asked if is plausible to make a migration from v1 to v2. Simon replied that it is undesirable, and it is better to re-process.

WGISS commented that at every meeting ARD gets clearer and clearer.

## Global Spatial Data Infrastructure (GSDI) Association

Gabor Remetey gave a report on the GSDI Association. He began with an update on selected activities of GSDI since WGISS-40. He listed recent GSDI conferences and their themes, and high-level cooperation with learned societies of the geospatial world.

Gabor highlighted the GEOSS foundational tasks where GSDI has the potential to contribute. These are in the areas of data sharing, developing the GCI, and community development and capacity building.

In terms of outreach, GSDI is interested in finding partners (WGISS involvement welcome) to contribute to the planned GSDI capacity building activities, which include the webinar programme that Dr. Joep Crompvoets and GSDI Secretary-General Roger Longhorn are putting together and workshops having SDI components. Webinars planned for 2016 and 2017 include those on SDI Policy and Legal Issues, SDI for Land Registration and Marine SDI Best Practice.

Gabor listed GSDI information sources, and discussed several conferences. He also listed relevant news from Europe, including NASA World Wind Europa Challenge 2016.

In news from Hungary, one highlight is the creation of an open source, cross platform, easy to use geospatial ecosystem intended for end users where selected/created data are stored in small raster and vector files and drawn using high performance render engine.

Gabor listed the following conclusions:

* Local, national, regional SDIs and associated services are enabling tools for EO applications and also vital from the local and regional Data Cube development point of view.
* GSDI is capable not only serving and supporting EO applications, but also capacity building, providing awareness raising and generating user feedbacks with special emphasis on user requirements.
* As GEOSS foundational tasks are concerned, GSDI is continuing support the GEOSS data sharing policy and ready to contribute to the data management principles implementation guideline development.
* The theme ‘Smart Homeland’ of coming GSDI World Conference in Taipei are matching many of the GA’s focused areas but having EO session and workshop possibilities it offers excellent outreach opportunity for the WGISS community.

Andy observed that the themes are very similar to those of WGISS and asked what WGISS can do to support that. Gabor said it would be very effective if a core member of WGISS could attend one of the conferences. Andy replied that if no one can attend, a report can be provided for Gabor to present.

Richard wondered if GSDI participates in the DMP Task Force; perhaps one of the GSDI members may be able to participate.

**Action WGISS-41-09e:** WGISS to consider contributions to inform the GSDI community about WGISS achievements.

## Indian Space Research Organization (ISRO)

Nitant Dube gave a presentation on the ISRO Indian EO Programme. He described this constellation of satellites (Resourcesat, Cartosat, and OceanSat), which is supported by strong ground mechanisms. The goals of the programme are to provide continuity of data with enhanced capabilities, to conduct periodic natural resources inventory and enable the national spatial data infrastructure, to maximize outreach of natural resources information, and to enable ocean state forecasting and improved weather forecasting.

Nitant discussed the evolution of missions, now in the third decade of processing EO data and providing it to users, and listed the currently operational and future EO missions. The future of EO missions will have applications in weather and ocean forecasting, natural resources monitoring, cartography, urban planning, disasters, and other land and ocean applications. In greater detail he described dual frequency radar imaging satellite to explore newer application areas using L and S band microwave data, as well as the National Natural Resources Management System, and the Disaster Management Support System.

Space applications are recognized in diversified areas, including agriculture, drinking water, fisheries, watershed development, monitoring irrigation infrastructure, and NR census, forestry, snow and glaciers, national urban information system.

The EO information services is a unique gateway to Indian EO data and services (Bhuvan), and the Meteorological and Oceanographic Satellite Data Archival Centre (MOSDAC).

Yonsook asked if the data from their upcoming missions will be accessible from their two data centres; Nitant replied that they will be as soon as they register the DIFs.

## Japan Aerospace Exploration Agency (JAXA)

Satoko Miura gave a report on JAXA, showing current and future satellite and sensor activities. She showed the time series of monthly mean CO2 results from GOSAT since 2008, and an image of the JAXA Real-time Rainfall Watch from the Himawari-8 satellite, which is now available. Satoko reported that the JERS-1 data archive is now available on the G-Portal. She also described the Forest Monitoring System that uses ALOS-2, with a resolution of 50 meters; the findings of the monitoring system will be updated about every six weeks on average to the JAXA website and other resources by March 2017.

## National Aeronautics and Space Administration (NASA)

Andy Mitchell gave a report about NASA’s Earth Science Data System (ESDS) Program, whose vision is to make NASA’s free and open Earth science data interactive, interoperable and accessible for research and societal benefit today and tomorrow. He listed the NASA ESDS priorities and space missions. Andy stated that the Earth Science Data System Program oversees the lifecycle of Earth science data with the principle goal of maximizing the scientific return from NASA’s missions and experiments for research and applied scientists, decision makers and society at large.

The data responsibilities and the goals of ESDS, which are required and funded, are:

* Set the standard for the archival and delivery of science-quality data related to planet Earth.
* Provide consistent harmonized services for data, metadata and services to facilitate interdisciplinary Earth science investigations.
* Ensure data is documented precisely and fully, and access to data and services are useful and useable by a wide community of users.
* Improve data system capabilities through continual evolution and community engagement.

A plan for prototypes is underway to evaluate advantages, risks, and costs of using commercial cloud technology for core Distributed Active Archive Center (DAAC) functions and this has completely shifted the focus of ESDS. Three prototypes are a web object storage and distribution prototype, an Advanced Rapid Imaging and Analysis (ARIA) processing demonstration prototype, and an ingest/archive prototype.

Hosting applications in the cloud offers the clearest path to improve efficiency. Over the next two years core applications will be adapted to work within commercial cloud environments. These include Metadata Search and Order System, Visualization System, Earthdata Web Infrastructure, and Earthdata Login. Core applications provide a foundation for future science discipline-specific applications to be developed on a shared, common architecture, as well as the opportunity to build new applications that leverage the shared services provided by the platform.

Mirko suggested a session on experiences and lessons learned about hosting applications in the cloud.

## National Oceanic and Atmospheric Administration (NOAA)

Martin Yapur gave a presentation on NOAA, who is working on developing intelligent architectures to satisfy the need for EO by analyzing the requirements early in the process. The value tree of data continues to grow, as does the volume of data. NOAA is largely data-driven organization; the environmental data growth is mostly from new satellites. Martin listed a number of drivers for change in the organization, toward the goal of developing a space-based observing enterprise that is flexible, responsive to evolving technologies, and economically sustainable.

To reduce complexity, NOAA is reorganizing its data centers. The idea is to have one consolidated data center, with the mission to be the stewards of the nation’s environmental information. This NOAA National Centres for Environmental Information (NCEI) distributes data stewardship responsibilities across the organization through Tiers of Data Stewardship. The Data Stewardship Division is responsible for the first two tiers, which involve long-term preservation through access services.

In summary, NOAA’s three data centers have been merged for about a year; the change has been challenging, but the opportunities for improved data stewardship are endless. Through the programs that are being developed, NCEI will significantly improve the long-term discoverability, accessibility and usability of NOAA’s data.

Mirko observed that the Data Stewardship Maturity Model would provide insights, and Andy noted that what NOAA is doing in terms of cloud computing and big Earth data is very interesting, and would be good to hear more about it.

## US Geological Survey (USGS)

Kristi Kline gave a report on USGS. Current major activities include Landsat Archive/Distribution Changes, Sentinel-2, the GloVis upgrade, and Landsat 9. She displayed the USGS product roadmap adding that USGS defined three basic categories of products: NRT, Tier-1 (which meet the criteria for the collection definition), and Tier 2 (which do not, but have been processed using the best known ancillary data). She gave further details of each of these.

Sentinel-2 is now available through Earth Explorer. USGS is providing individual tiles, and created a 3-band full resolution browse. Three interfaces are EarthExplorer, GloVis, and LandsatLook.

GloVis is a quick and easy online search and order tool. It is old and has a lot of problems, so it is being totally redesigned using modern languages and image processing tools. She displayed the conceptual architecture.

Kristi explained the role of Landsat, and explained the relationship between NASA and USGS. She described the Landsat 9 mission objectives, and development approach. She also displayed the Landsat reception network, and showed the project schedule milestones. Landsat 9 is expected to launch in 2020.

# WGISS Plenary, Part II

## Future Meetings

Mirko Albani announced that WGISS-42 is scheduled for September 19-23 in Frascati, Italy, hosted by ESA-ESRIN. He described the location, method of arrival, and the available hotels in Frascati. He also noted that participants can stay in Rome. Transfers from Frascati to ESRIN will be provided. Mirko described the social events planned, and a tour of the ESRIN facility.

The preliminary plan for the meeting is three full days of plenary meeting, with Thursday and Friday morning for an FDA workshop. Andy suggested that the Technology Exploration session be on Wednesday including presentations on FDA, as a precursor to the workshop.

It will be necessary to develop an invitation, identify the sessions and potential participants, and to specify what parts of ARD are important so the meeting is focused. Participation may be limited to 60-70. Feedback from SIT will be obtained.

Other future meetings.

WGISS-43: Possibly April 3-7, 2017 in the Americas, subject to SIT approval. Possible hosts are NASA or NOAA.

WGISS-44: September 2017 in Asia or Africa. Mirko listed possible hosting options.

**Action WGISS-41-13a:** WGISS Exec to work on identifying objectives and target audience for the FDA Workshop and present these plans to the SIT at the April meeting.

## Chair Summary

Andy summarized this very productive meeting with the following highlights:

1. Collaboration with WGCV:
	1. WGISS and WGCV will present jointly at the 31st CEOS SIT Meeting (April 19-20) on how the two working groups are performing an analysis of the Carbon Action Items in order sufficiently address them.
	2. Other possible joint activities include:
		1. Support Metadata Requirements for Quality (Quality fields)
			1. Review and comment on GEOSS Data Quality Requirements (DDQ -Documentation on Data Quality)
			2. Include validation metrics (possibly calibration metrics and sensor performance indicator) Make the reference to validation guidelines into the metadata information
			3. Develop a White Paper on Recommendations for improving Data Quality in Metadata
		2. Develop Best Practices on Data Reprocessing – including current and future practices (i.e. on-demand ARD Processing)
2. CEOS Work Plan. WGISS has an action to give feedback on section 4 of the document; this will be a section on WGISS efforts on Big Data and Cloud Computing.
3. ARD: WGISS will work with SEO, WGCV and LSI on the ARD definition and on enabling discovery of ARD to support Data Cube implementations. WGISS will also provide guidelines for the metadata for ARD to ensure the information needed to describe a measurement is available, focusing on GFOI and GEOGLAM.
4. Additional WGISS support needed for Foundational tasks GD-03, GD-05, and GD-08. WGISS is already tasked to be involved in GD-02 and GD-07.
5. Current architecture for WGISS assets: Currently there are multiple entry points; the goal is to present a unified architecture. WGISS will begin this conversation.
6. A list of Technology Exploration webinars was developed, and work on these will begin.
7. Additional new initiatives:
	1. Thesaurus (EO keywords)
	2. Interoperability of metrics (This would be extremely valuable to CEOS), user requirements, users profiles (ties into metrics conversation), unified metadata model
	3. Metadata catalog that supports the CEOS Handbook. For better uniformity.

**Action WGISS-41-06d: Interoperability of metrics, user requirements, user profiles, unified metadata model.**

**Action WGISS-41-06e: Metadata catalog that supports the CEOS Handbook.**

## WGISS-41 **Actions**

The actions resulting from WGISS-41 are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sub-task** | **Category** | **Description** | **Actionee** |
| WGISS-41-01 | Carbon | WGISS Exec to present jointly with WGCV at the 31st CEOS SIT Meeting (April 19-20) how the two WGs are performing an analysis of the Carbon Action items in order to sufficiently address them. | WGISS-Exec |
| WGISS-41-02 | Data Quality | In order to support, with WGCV, Metadata Requirements for Quality (Quality fields), confirm with Albrecht if these are activities we should pursue: | Andy Mitchell |
| WGISS-41-02a | Data Quality | Review and comment on GEOSS Data Quality Requirements (DDQ -Documentation on Data Quality) | Andy Mitchell |
| WGISS-41-02b | Data Quality | Include validation metrics (possibly calibration metrics and sensor performance indicator). Include the reference to validation guidelines into the metadata information. | Andy Mitchell |
| WGISS-41-02c | Data Quality | Develop a White Paper on Recommendations for improving Data Quality in Metadata. | Andy Mitchell |
| WGISS-41-03 | FDA | Confirm with Albrecht the possibility to work with WGCV to develop Best Practices on Data Reprocessing – including current and future practices (i.e. on-demand ARD Processing). | Andy Mitchell |
| WGISS-41-04 | FDA | WGISS to review relevant CEOS and agency activities and ambitions in order to support the study on future data architectures: WGISS efforts on Big Data and Cloud Computing; and Provide Technology Exploration Interest Group summary on past efforts. |   |
| WGISS-41-05 | FDA | Richard and Mirko to organize telecon with RDA on the best practice and recommendations for DOI/PI. | Richard Moreno, Mirko Albani |
| WGISS-41-06 | New Initiatives | WGISS to work on the following new initiatives: |   |
| WGISS-41-06a | New Initiatives | A Unified Architecture for WGISS Assets. | Yonsook Enloe, Mirko Albani |
| WGISS-41-06b | New Initiatives | To discuss the concept of developing and sustaining a WGISS discovery and access system level team. | Yonsook Enloe, Mirko Albani |
| WGISS-41-06c | New Initiatives | A Thesaurus (EO keywords) | Michael Morahan, Richard Moreno |
| WGISS-41-06d | New Initiatives | Interoperability of metrics, user requirements, user profiles, unified metadata model. |   |
| WGISS-41-06e | New Initiatives | Metadata catalog that supports the CEOS Handbook. |   |
| WGISS-41-07 | Outreach | WGISS to develop monthly webinars; several topics have been suggested and should be prioritized. | Technology Exploration IG |
| WGISS-41-08 | Outreach | In order to discover how WGISS can assess the needs of other CEOS working groups, virtual constellations, and ad-hoc groups with regards to WGISS capabilities: | WGISS-Exec |
| WGISS-41-08a | Outreach | WGISS members to identify, from within their own agencies, members of said groups. | WGISS members |
| WGISS-41-08b | Outreach | WGISS-Exec to coordinate one-on-one interactions with said contacts. | WGISS-Exec |
| WGISS-41-09 | Outreach | In order to improve WGISS outreach: |   |
| WGISS-41-09a | Outreach | WGISS-Exec to investigate ways to develop additional outreach materials. | WGISS-Exec |
| WGISS-41-09b | Outreach | WGISS Exec to start planning personal engagement for invitations to future meetings. | WGISS-Exec |
| WGISS-41-09c | Outreach | WGISS Exec to put together a list of the benefits and capabilities of WGISS to assist the conversation. This list will also be put on the WGISS website. | WGISS-Exec |
| WGISS-41-09d | Outreach | WGISS-Exec to provide a news item for the CEOS webpage using the Recovery Observatory as a story of good collaboration between working groups. | WGISS-Exec |
| WGISS-41-09e | Outreach | WGISS to consider contributions to inform the GSDI community about WGISS achievements. | WGISS-Exec |
| WGISS-41-10 | GEO | WGISS Exec to review: | WGISS-Exec |
| WGISS-41-10a | GEO | GEO Foundational Task GD-03 to identify where WGISS can contribute. | WGISS-Exec |
| WGISS-41-10b | GEO | GEO Foundational Task GD-08 to identify where WGISS can contribute, and propose this to the CEO, as it is linked with G05. | WGISS-Exec |
| WGISS-41-11 | GEO | WGISS Exec to suggest to CEO: | WGISS-Exec |
| WGISS-41-11a | GEO | To merge GEO Foundational Subtask GD07.05 into GD07.01. | WGISS-Exec |
| WGISS-41-11b | GEO | To remove GEO Foundational Subtask GD07.03 from GEO Foundational Task GD07, and make it a separate Foundational Task or move into GD01. | WGISS-Exec |
| WGISS-41-12 | GEO | Andy and Martin to check to see if the USGEO Data Management Principles information is feeding to GEO. | Andy Mitchell, Martin Yapur |
| WGISS-41-13 | WGISS-42 | WGISS Exec to include the following in the planning for WGISS-42: | WGISS-Exec |
| WGISS-41-13a | WGISS-42 | WGISS Exec to work on identifying objectives and target audience for the FDA Workshop and present these plans to the SIT at the April meeting. | WGISS-Exec |
| WGISS-41-13b | WGISS-42 | Andy to give a presentation at WGISS-42 on what NASA is doing with the UMM to standardize the metadata model. | Andy Mitchell |
| WGISS-41-13c | WGISS-42 | DSIG to give a presentation on the investigation of a CEOS-recommended metadata model. | DSIG |
| WGISS-41-14 | MIM | Regarding the MIM: |   |
| WGISS-41-14a | MIM | Michelle to send to agency representatives instructions about updating the MIM spreadsheet with URLs to the mission website and a URL to the data access portal for the mission data. | Michelle Piepgrass |
| WGISS-41-14b | MIM | Agency representatives to populate the MIM with their changes to the mission website URL and the mission data access portal URL. | Agency Representatives |
| WGISS-41-14c | MIM | Andy to follow up with the MIM team about suggestions to enhance the MIM and give an update at WGISS-42. | Andy Mitchell |
| WGISS-41-14d | MIM | Brian to talk with Michael, Yonsook and Mirko about the unique ID for mission and instrument combination in CWIC and FedEO. | Brian Killough |
| WGISS-41-15 | Website | WGISS Exec to work on the organization of the Documents web page and content of documents. | WGISS-Exec |

## Adjourn

Andy adjourned the meeting, thanking the participants for their contribution to a productive and successful meeting. He also thanked GA and CSIRO, who presented the intriguing work they have done, are doing, and want to do in future. WGISS looks forward to their support, as there is an explosion of work for the future that is very exciting and will prove to be very useful.

# Glossary of Acronyms

API Application Programming Interface

CEO CEOS Executive Officer

CEOS Committee on Earth Observation Satellites

COTS Commercial Off-the-Shelf

CSW Catalogue Service for the Web

CWIC CEOS WGISS Integrated Catalogue

DAAC Distributed Active Archive Center

DC Data Cube

DIF Directory Interchange Format

ECV Essential Climate Variable

EO Earth Observation

ESIP Federation of Earth Science Information Partners

GCI GEOSS Common Infrastructure

GEO Group on Earth Observations

GEO-GLAM Global Agricultural Monitoring

GEOSS Global Earth Observation System of Systems

GFOI Global Forest Observations Initiative

GIS Geospatial Information System

GPM Global Precipitation Mission

GPU Graphics Processing Unit

GSDI Global Spatial Data Infrastructure

GUI Graphical User Interface

HPC High Performance Computing

IDN International Directory Network

ISO International Standards Organisation

LSI Land Surface Imaging

NRT Near real-time

OGC Open Geospatial Consortium

PI Persistent Identifier

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