WGCV/WGISS interactions

Richard MORENO
WGISS Chair
CNES

WGISS ?

- **Working Group on Information Systems and Services.**
  - **A subsidiary body supporting the Committee on Earth Observing Satellites (CEOS).**
  - **Promotes collaboration in the development of systems and services that manage and supply these observatory data.**
  - **Creates and demonstrates prototypes supporting CEOS and Group on Earth Observation (GEO) requirements.**

WGISS ?

- **Working Group on Information Systems and Services.**
  - **Addresses**
    - **the internal management of EO data,**
    - **the creation of information systems**
    - **and the delivery of interoperable services.**
  - **The activities and expertise of WGISS span the full range of the information life cycle**
    - **from the requirements and metadata definition for the initial ingestion of satellite data into archives**
    - **through to the incorporation of derived information into end-user applications.**

WGCV37
WGCV/WGISS interactions
20 February 2014

WGISS Structure

WGISS
Richard Moreno, CNES, Chair
Andrew E. Mitchell, NASA, Vice Chair

WGISS Executive Secretariat
Michelle Piepgrass, CNES

Interest Groups (IG)	
Technology Exploration	**Andrew Mitchell, NASA**
Data Stewardship	~~John Faundeen, USGS~~
Intl. Directory Network (IDN)	**Lola Olsen, NASA**
Virtual Constellation	~~John Faundeen, USGS~~
Projects (Prj.)	
WGISS Infrastructure Services	**Martin Yapur, NOAA**
CWIC	**Martin Yapur, NOAA, Yonsook Enloe, NASA**
~~**GA.4.Disasters**~~	~~**Karen Moe, NASA, Sergii Skakun, NSAU**~~
Water Portal	**Satoko Miura, JAXA**
CEOS OpenSearch	**Jérôme Gasperi, CNES Yoshiyuki Kudo, JAXA**

Technology Exploration IG

- A community forum for WGISS members related to emerging technologies !
- The goal is to allow members who are working on similar technologies to learn from each other's experiences.
  - Members will be able to share real-world (success and failure) stories or case studies about how their systems have approached implementation and the effects to their communities.
- 2014 subject areas will include (but not limited to)
  - Authentication Services (SSO, …),
  - Big Data,
  - Semantic search
  - Cloud Computing.

Quality

Data Stewardship IG

- **Goals:**
  - **Enable the sharing of agency investigations, developments, and lessons-learned relating to EO data stewardship**
  - **Share experiences and lessons learned**
  - **Draft common cross-agency best practices or guidelines of data stewardship for possible adoption by WGISS**
  - **Sponsor technical exchanges at WGISS meetings or elsewhere**

Quality

Data Stewardship IG

- **Scope: Focus on Data, Metadata and Product Topics Including:**
  - **Long-Term Archive Strategies**
  - **Data Formats**
  - **Data Preservation**
  - **Data Lifecycle Concepts**
  - **Archive Media**

Virtual Constellation

- New VC liaison position to assist WGISS with establishing a close working relationship with VCs
- CNES's Data Access Study is under review
  - All the datasets mentioned in this study are hosted by agencies that are currently WGISS members.
  - WGISS's priority will be working on getting these datasets published in the IDN and accessible via Opensearch/CWIC
  - The Land Surface Imagining (LSI) Portal, Atmospheric Composition Constellation (ACC) Portal and the JAXA Water Portal are heavily discussed and reviewed in WGISS

IDN Interest Group

Data Discovery

- http://idn.ceos.org – International Directory Network
- An international effort developed to assist users in locating data sets, data services, and visualizations.
- Provide free, online access to metadata on scientific data in the Earth sciences: geoscience, hydrospheric, biospheric, satellite remote sensing, and atmospheric sciences.
- Provide clients/portals directory-level search of CEOS agencies and others who have registered their collections in the IDN

Quality

Data Discovery

IDN Interest Group

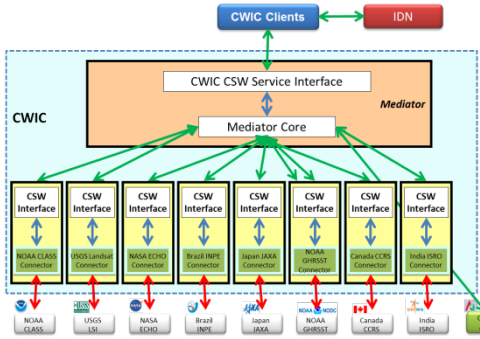
- IDN is harvested by the GEOSS/GEO
- IDN web services are harvested by various data brokers
- IDN contains link to interoperable data access points
  - CWIC or opensearch
  - Allows to seamlessly discover and access the data

- It is very important that CEOS agencies update their metadata in IDN

Data Access

CWIC Project

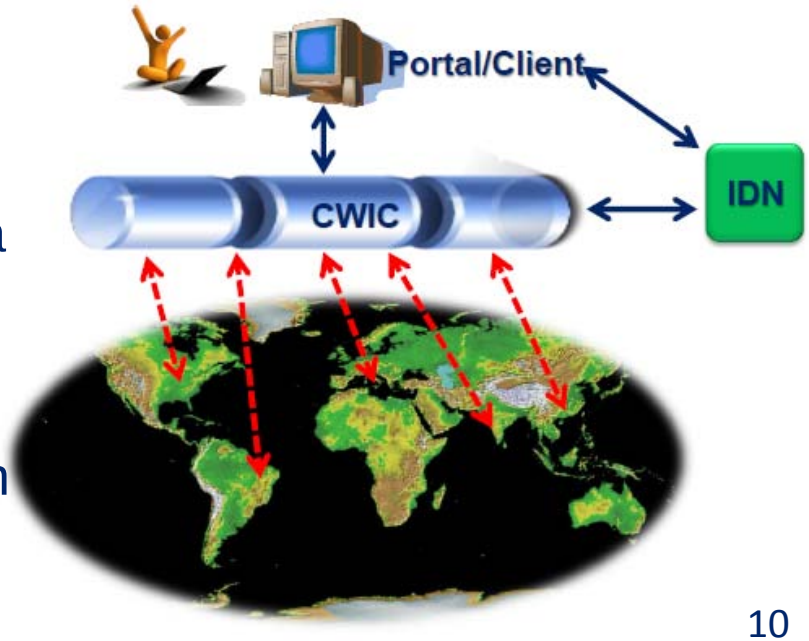
- **CWIC = CEOS WGISS Integrated Catalog**
- **Implemented to realize a federated catalogue for data discovery from multiple EO data centers**
- **The CWIC architecture consists of 3 main items:**



•CWIC middleware - provides an access point for an inventory-level search at CEOS agencies who are CWIC data providers

•CWIC clients/portals - user interfaces to access cross-discipline data from CWIC data providers

•IDN – International Directory Network (IDN) provides clients/portals directory-level search of CEOS agencies who have registered their collections in the IDN



Data Access

CWIC Dataset registration Status

(1) Numeric (21 Jan 2014)

- # of discoverable and accessible collections = 1841
- # of discoverable and accessible granules >67 millions

(2) CWIC Data Partners

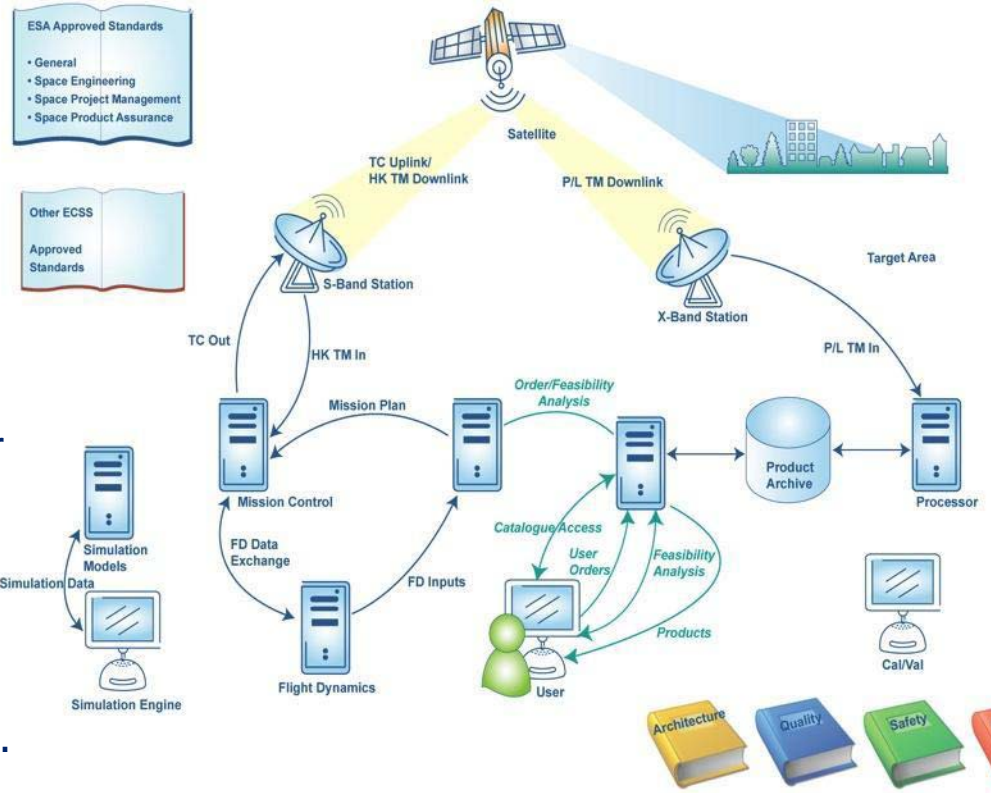
- Operational : NASA, USGS, LSI, INPE, SST VC (GHRSST/NODC), CCMEO (formerly CCRS)
- Still in TEST: AOE, ISRO, NOAA CLASS
- Will be in TEST: ROSCOSMOS

Data Access

What is HMA

A collaborative project started in Europe and Canada by the Ground Segment Coordination Body (GSCB) in 2005 with the objective to:

- Guarantee a seamless and **harmonised access to heterogeneous EO datasets from multiple mission** ground segments, including national missions and ESA missions.
- **Standardise** the ground segment **interfaces** of the satellite missions for easier access to EO data.
- Provide **interoperability for coordinated data access** enabling the interactions with services or Value Adders and EO Contributing Missions.



Data Access

FedEO Dataset registration Status

(1) Numeric

- # of discoverable collections = 433
- # of accessible collections = 433
- # of discoverable granules > 6 millions

Data Access

FedEO-CWIC Interoperabitliy

CEOS portals

GEO portal

OpenSearch compliant missions
(i.e. no CWIC, no FedEO)

IDN

CWIC

OpenSearch

FedEO

CWIC

CWIC missions

NOAA
CLASS
GHRSST

USGS
LSI

ISRO

NASA

CCMEO

AOE
Beijing-1
NSMC/FengYun
CRESDA

INPE

OpenSearch

CNES

Disasters
Charter

HMA

OpenSearch

ESA / ESRIN

EO-DAIL

ESA
missions

Other
missions

NASA
ECHO
Gateway

HMA

ECHO
specific

European missions

DLR

EUMETSAT

SPOT
IMAGE

VITO

DMC

MDA

NASA

NASA
ECHO

FedEO CWIC Interoperability

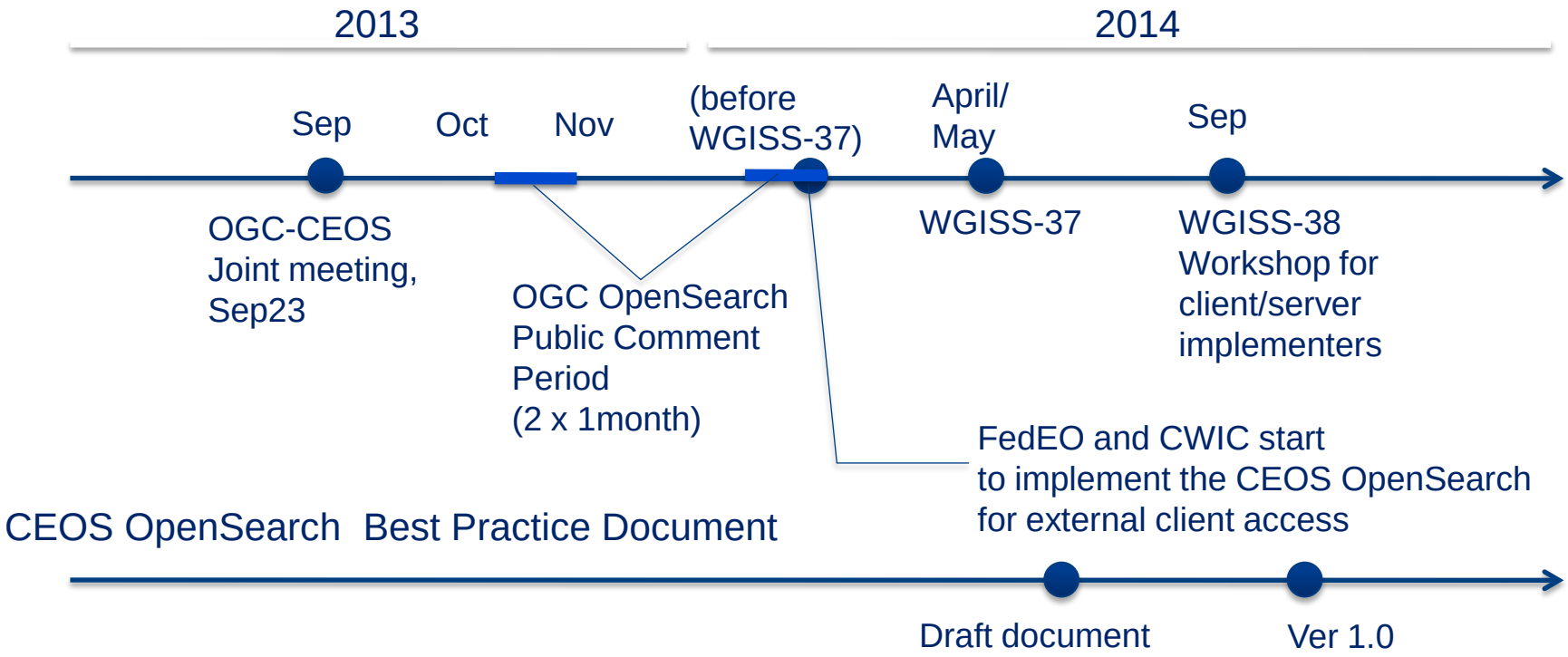
OGC CSW OGC 10-032 OpenSearch OGC 06-131 EO AP (HMA) ECHO specific

Easy data Access

CEOS OpenSearch Project

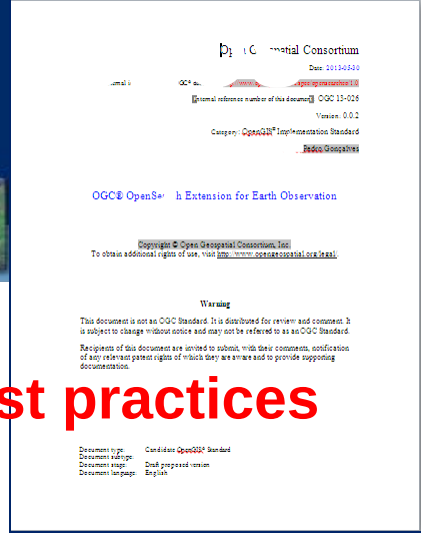
"CEOS OpenSearch"

- Establishes a common CEOS interoperability best practice of OpenSearch to allow standardized and harmonized access to metadata and data of CEOS member agencies, including CWIC and FedEO communities



Easy data Access

OpenSearch



- **Opensearch standard = 2 documents**
  + CEOS Opensearch Best practices

- **The 1st document : "OpenSearch GeoSpatial and Temporal Extensions" OGC 10 032r6**
  - **Comment phase is now closed ; and WGISS comments have been taken into account**
  - **A "CEOS Best practice document" has been issued and is being discussed**

- **The 2nd document : "OpenSearch Extension for Earth Observation" OGC 13 026**
  - **OGC comment phase will start in OGC TC meeting in March**
  - **The "CEOS Best practice document" will be updated**

Quality

CEOS Water Portal

- A distributed data system component of DIAS (Data Integrated Analysis System)-Program
- To provide “Easy to Access” service to users
- To provide access to a whole variety of hydrological data and water relevant data scattered over the world (incl. Satellite data, In-situ data and Model data)
- To connect the existing components like data centers, scientists and wide users.

More than 2 millions data files can be accessed from this portal.

http://waterportal.ceos.org/

Next WGISS meeting

- WGISS-37 will be held in Cocoa Beach, Florida, US, April 14-18, 2014.

- The meeting will be hosted by NASA

WGCV / WGISS interactions

WGCV – WGISS interactions

WGCV / WGISS interactions

Why do WGISS need cooperation with WGCV ?

- Thanks to interoperability progress, more and more datasets (several hundreds) and data (several dozens of millions) are available from a single point or from any standard access tool

- Data Ranking is becoming crucial
  - Via semantics,
  - Via metadata enrichement
    - Quality information is one of the most important
    - Shall be standardized and included in interoperability standards : Opensearch

WGCV / WGISS interactions

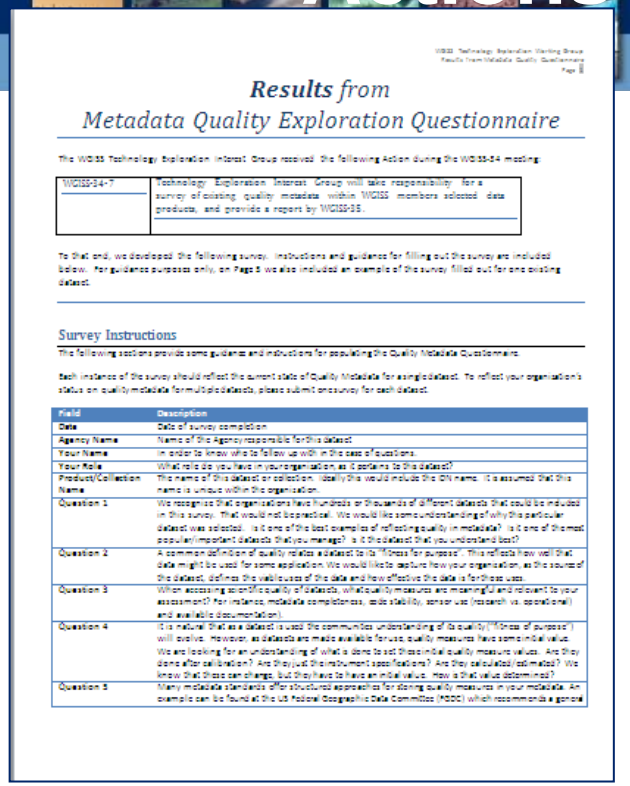
Why do WGISS need cooperation with WGCV ?

- Digital data is perishable
- Data quality information is essential for data preservation
  - Data long term preservation is expensive
    - Quality information can help in the decision of preserving or not a datasets or a data
  - The data quality is a mandatory metadata
- Data preservation needs standards to ensure the long term usability

WGCV/WGISS Actions

Summary

1. WGCV teams should also suggest ideas of quality data metadata fields for the key sensors and products.
   - WGISS to look at quality metadata for WGISS 35, starting from collection level (led by Tech Expo IG).
   - Common definitions on quality is needed. WGCV will work on this.
   - Work together based on above results and prepare a example(s). Show and appeal to CEOS members !



***Results** from*
Metadata Quality Exploration Questionnaire

The WGISS Technology Exploration Interest Group received the following Action during the WGISS-34 meeting:

WGISS-34-7	Technology Exploration Interest Group will take responsibility for a survey of existing quality metadata within WGISS members selected data products, and provide a report by WGISS-35.

To that end, we developed the following survey. Instructions and guidance for filling out the survey are included below. For guidance purposes only, on Page 6 we also included an example of the survey filled out for one existing dataset.

Survey Instructions

The following sections provide some guidance and instructions for populating the Quality Metadata Questionnaire.

Each instance of the survey should reflect the current state of Quality Metadata for a single dataset. To reflect your organization's status on quality metadata for multiple datasets, please submit one survey for each dataset.

Field	Description
Date	Date of survey completion
Agency Name	Name of the Agency responsible for this dataset
Your Name	In order to know who to follow up with in the case of questions
Your Role	What role do you have in your organization, as it pertains to this dataset?
Product/Collection Name	The name of this dataset or collection. Ideally this would include the GCMD name. It is assumed that this name is unique within the organization.
Question 1	We recognize that organizations have hundreds or thousands of different datasets that could be included in this survey. That would not be practical. We would like some understanding of why this particular dataset was selected. Is it one of the best examples of reflecting quality in metadata? Is it one of the most popular/important datasets that you manage? Is it the dataset that you understand best?
Question 2	A common definition of quality relates a dataset to its "fitness for purpose". This reflects how well that data might be used for some application. We would like to capture how your organization, as the source of this dataset, defines the value/success of this data and how effective this data is for those uses.
Question 3	When accessing specific quality of datasets, what quality measures are meaningful and relevant to your assessment? For instance, metadata completeness, code stability, sensor use (research vs. operational) and available documentation).
Question 4	It is natural that as a dataset is used the communities understanding of its quality ("fitness of purpose") will evolve. However, as datasets are made available for use, quality measures have some initial value. We are looking for an understanding of what is done to set those initial quality measure values. Are they done after calibration? Are they just the instrument specifications? Are they calculated/estimated? We know that these can change, but they have to have an initial value. How is that value determined?
Question 5	Many metadata standards offer structured approaches for storing quality measures in your metadata. An example can be found at the US Federal Geographic Data Committee (FGDC) which recommends a general

- **WGISS conducted a survey in the form of a questionnaire of existing quality metadata within our member data products.**
- **Results of that survey represent responses from JAXA, INPE, and a sample provided by NOAA.**
- **If desired, WGISS is willing to continue the discussion of quality metadata within WGISS.**

WGCV/WGISS Actions

2. Metadata requirements for quality, need to tap each WGCV working group for sensor information. Have WGISS find out what is available. Get NASA ESIP feed back on this and others.
   - No action
   - WGCV & WGISS chairs will send e-mail WGISS/WGCV members to document all the quality related activities.
   - Nigel will provide QA4EO

Related to #1.

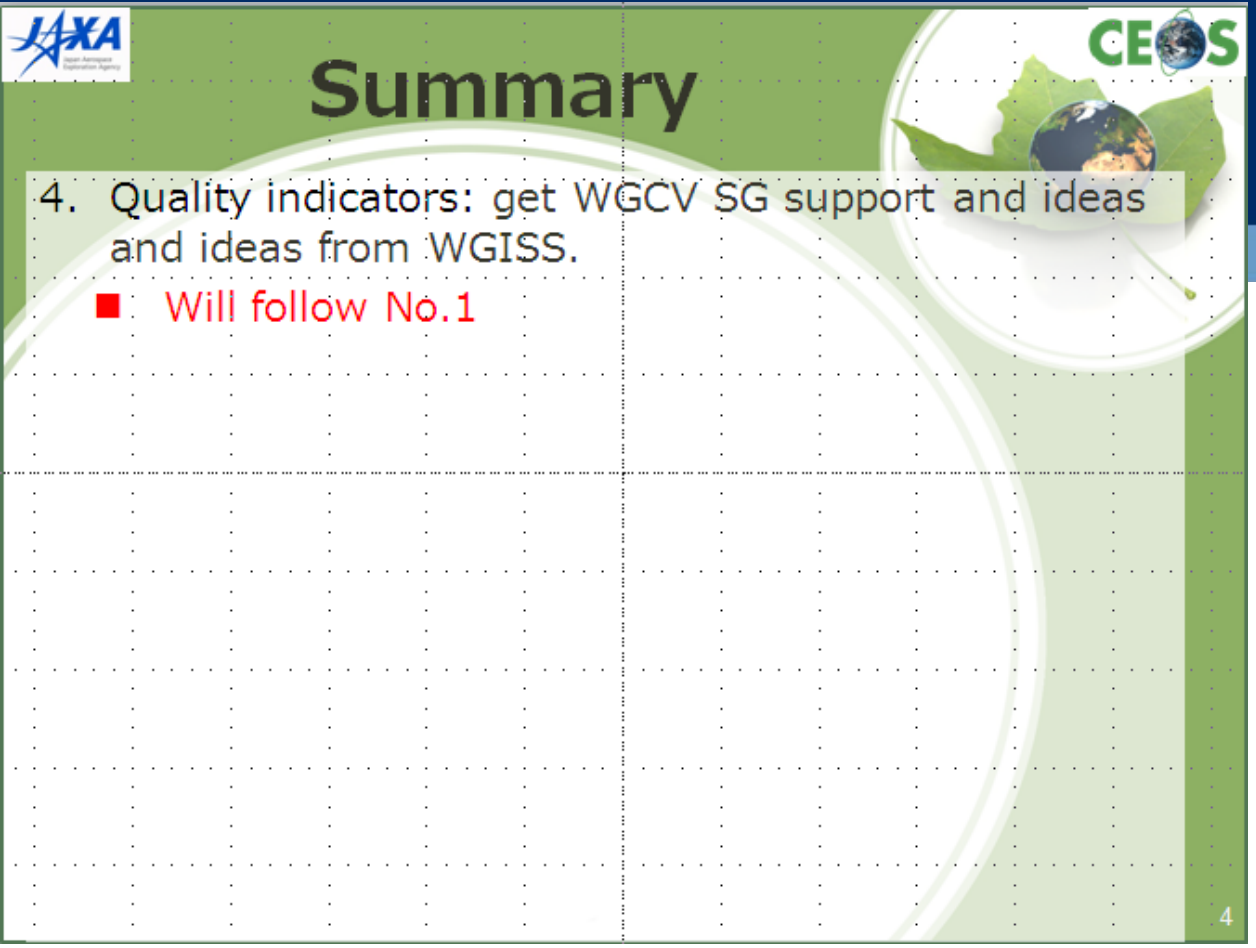
- **WGISS and WGCV have been invited to participate in the GEO Data Sharing Working Group-Data Documentation and Quality (DDQ) Subgroup. The DDQ subgroup prepared the GEOSS Data Quality Guidelines which is meant to be a general overview document that points users in the right direction of data quality information.**
- **Also, NASA's ESDSWG (Earth Science Data Systems Working Group) is forming a committee to assess current data quality related activities in the inter-agency/international arena and recommend how to apply them for products from NASA Earth science data activities.**

WGCV/WGISS Actions

3. Data access of CEOS Test Site information starting with:
   a. CEOS IVOS sites (also test first one prior to meeting for Libya 4 site)
   b. LPV sites
   c. SAR
   d. DEM and others. Recommend starting with some key examples. SG support needed.

- WGCV will provide necessary information to WGISS/CWIC.
- CEOS WGISS Integrated Catalog (CWIC) will support for this.
- Data access to one test site will become available before the 2012 CEOS Plenary.

- **Some of the datasets provided by the USGS CWIC provider contain WGCV CEOS Test Sites data.**
- **CEOS opensearch srandard interface is very easy to implement**



- **Maybe an input for “OpenSearch Extension for Earth Observation” OGC 13 026**

WGCV/WGISS Actions

Summary

5. How does quality assurance information on ECV products get populated and stored and what are the documentation components needed
   - WGCV will contact to Climate WG to obtain quality assurance requirements.
   - WGISS/DSIG will contribute to this, especially regarding long term data preservation.
   - Input to GEO WP Component IN-02-C1

- **Maybe an input for “OpenSearch Extension for Earth Observation” OGC 13 026**

WGCV/WGISS Actions

6. Identifying key partners and how they benefit the working group and they benefit from WG

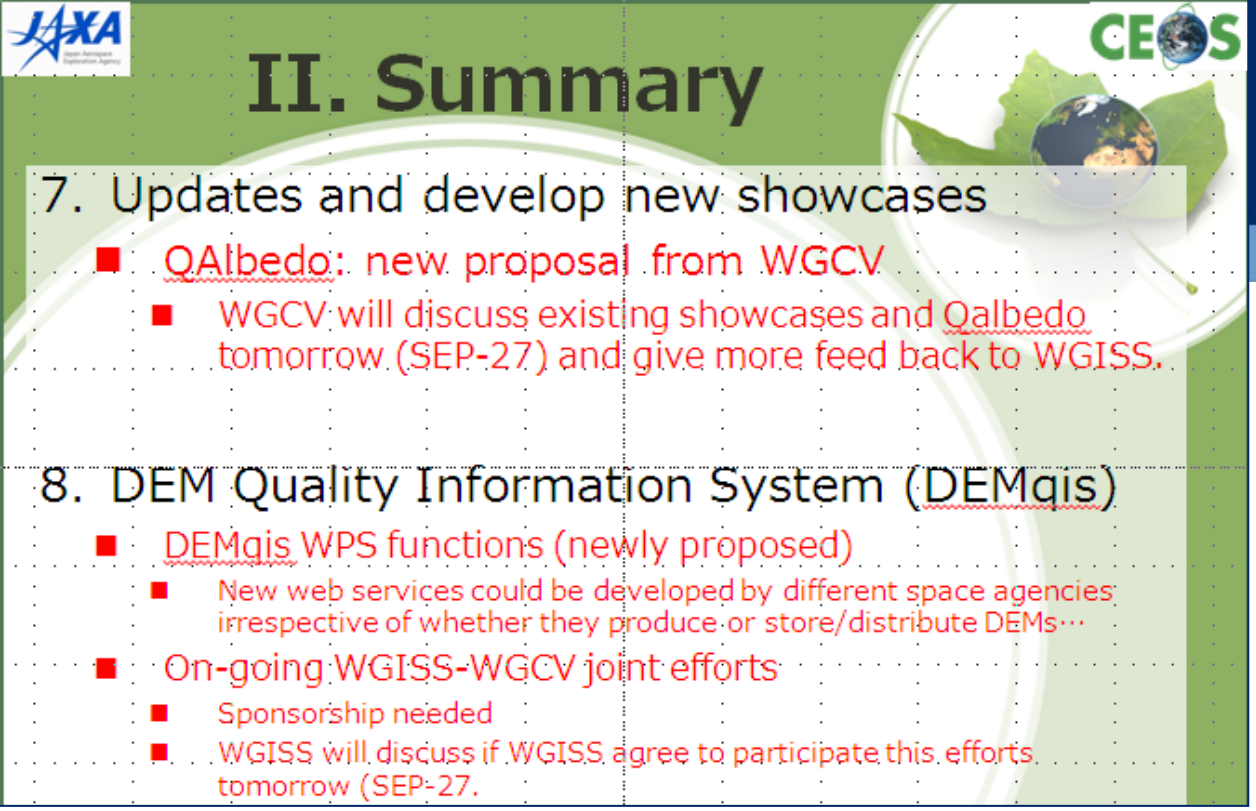
   - WGISS use Wish List from VCs. WGISS, working with WGCV , to contact to each VCs on their needs (detailed/specific needs).

- **WGISS is reviewing Steven Hosford's (CNES) presentation on VC's Data Access Study.**

- **All the datasets mentioned in his slides are hosted by agencies that are currently WGISS members.**

- **WGISS's priority will be working on getting these datasets published in the IDN and accessible via CWIC or/and Opensearch.**

WGCV/WGISS Actions



- **No update on QAlbedo. But in teleconference meetings (March-April, 2013) of Task Force on QA4EO Implementation, several potential case studies to demonstrate QA4EO principles have been discussed.**
  - **Examples are: TRUDAT, NOAA Maturity Matrix, ESA Sentinel-2 calibration tool, GSICS GEO-LEO, 3D Vegetation Lab and some work done under SST VC.**
  - **More discussions on deciding case studies to follow in next few meetings.**

Summary

9. Portal Quality and Commonality/new portals

10. Linking of CEOS tools and accessibility of tools

11. Data Accessibility and availability-list per mission? Data Policy-CAL/VAL-free and open

12. CEOS related calendar
    - WGCV chair to contact Kerry/DCEO if SG events (and other lower level events) can be reflected into the current calendar.

8

WGCV / WGISS interactions

- **Joint meetings**
  - 2008 in Sanya
  - 2010 in Montreal
  - 2012 in Hyderabad
  - Need / possibility for a joint meeting in 2015?

- **Side meetings during other CEOS meetings ?**
  - SIT meeting
  - SIT Workshop
  - CEOS plenary