CEOS/WGISS Technology & Services Annual Report
October 2006

The Technology and Services Subgroup (TSSG) of the CEOS Working Group on Information Systems and Services (WGISS) covers the internal management of data, the creation of information systems and the delivery of interoperable services.

The interests of the Subgroup span the full range of the information processing chain from the initial ingestion of satellite data into archives through to the incorporation of derived information into end–user applications. The creation and standardization of metadata to support the discovery of, and access to, data is an important aspect of this work.

The TSSG comprises six task teams, each one devoted to one of the topics covered by the subgroup :

> the Archive Task Team
>
> the International Directory Network task team
>
> the Interoperable Catalogue Systems task team
>
> the Data Services Task Team
>
> the Grid Task Team
>
> the EOGEO workshop

Summary report

It can be noted, since last CEOS plenary, through the work of the TSSG task teams, that there is now a strong trend to implement EO data systems as service oriented systems, with services defined and implemented according to the OGC and/or ISO recommendations, especially the OGC recommendations for web map services and the ISO recommendations for metadata. The Grid technology that has been analysed by the TSSG for several years is likely to become the underlying infrastructure technology for interoperable service oriented systems. In this perspective, the TSSG has started to explore the concept of "International Virtual Earth Observatory" and the concept of "Wide Area Grid". Operational systems like the IDN, the ICS eoPortal and ECHO are being continuously improved. They are the TSSG recommended solutions which could immediately fulfil some of the GEO requirements. The TSSG is also collecting recommendations and good practices in the field of data archiving since one of the greatest challenges for agencies will be to preserve and leave accessible to users their increasing amount of EO data over long periods of time. More generally, all TSSG recommendations and good practices for interoperability will be recorded in a "CEOS Interoperability Handbook".

Detailed report

The progress that has been made within each of these task teams since the last CEOS plenary is detailed hereafter in the following paragraphs.

Archive Task Team

1. Technology

The task team has been considering the future "SAFE" format currently being defined by ESA for the archiving of telemetry data. "SAFE" actually encapsulates the CCSDS XSFDU format.

USGS has released to WGISS a document that provides an assessment of the options for the next generation of off–line digital archive storage technology to be used for the Digital Archives of the USGS. The selected technology must be capable of safely retaining data until space, cost, and performance considerations drive the next migration. Data should be migrated before reliability degrades.

2. Data Preservation

The "purge alert" function is maintained by USGS on behalf of CEOS.

USGS has shared their experience in converting large amounts of data from tapes to digital formats (200 TB).

Also USGS has presented their long term archive strategy. A comparison between DAS (Direct Area Storage), NAS (Network Area Storage), SAN (Storage Area Network) technologies for long term archiving has been given.

International Directory Network Task Team

Version 9.6 of the Master Directory software is now available. A key feature of this version is the ability of requesting data on relative time periods. New thematic keywords for platforms and instruments have been added as "keyword hierarchies" (taxonomies) to the existing sets of controlled keywords. Instruments and platforms may now be grouped according to categories.

New IDN Portals have been added in response to CEOS/GEOSS requests.

The prototype of a multilingual IDN portal has been demonstrated, providing on the fly "Google" translations from English to French, Japanese and Thai languages.

Links to map services are available through the Related_URL_Content_Type in the metadata.

The IDN team is working on a profile of the ISO 19115 metadata standard. Additionally, there are efforts to combine the US and Canadian ISO 19115 profiles into a single North American profile for geospatial metadata.

Interoperable Catalogue System Task Team

1. ECHO

ECHO is the EOSDIS clearinghouse. It is an open system developed by NASA that provides Earth science data and services to large, diverse pool of users enabling scientific community interaction and collaboration. Today, it allows metadata ingestion, data search (spatial search, temporal search, keywords search, numeric search types), data access (online, order brokering). Web service based open interfaces are available for human–machine or

machine–to–machine clients. CEOP, JAXA, the Israel Space Agency, IRE RAS, ESA and Dundee University are using or considering interfacing to or using ECHO.

2. HMA

"Heterogeneous Mission Accessibility" (HMA) is a project being lead by ESA in cooperation with European agencies (ASI, CNES, DLR...) which should enable the interoperable access of optical, radar and atmospheric space borne data in Europe. The protocols developed within HMA are intended to be aligned as implementation profiles of the OGC recommendations for data cataloguing and data searching. Searches will be distributed to various catalogues through a service orchestration mechanism.

3. Ontologies applied to Catalogues

An ontology is a formalization for representation of knowledge based on the mathematical theory of Descriptive Logics. CNES is adding an experimental extension to his metadata system that will allow "ontology driven" queries going to the metadata database. The objective of this extension is to get greater flexibility for the definition of these queries (current queries only allow for space–time and keyword based selections).

4. CIP

The Catalogue Interoperability Protocol (CIP) is a Z39.50 based protocol for the interoperable access to data. CIP was defined by CEOS and has been prototyped with the INFEO system, now the eoPortal system maintained by ESA. eoPortal is an operational system. However, the CIP specification is now frozen and no additional changes are foreseen.

5. OGC catalogue

The ICS task team is continuously analysing the OGC recommendations for data catalogues. In particular following the evolutions in OGC where currently there is a proposal that ebRIM may become the only metamodel for OGC compliant catalogue implementations.

Data Services Task Team

1. IDN Services

The IDN started to provide service metadata (know as "Service Entry Resource Format" – SERFs) in 1999. More than 1500 SERFS are now available. They provide descriptions and direct links to Earth science tools and software, including GIS software, models, educational resources, hazards management, data/metadata management, and environmental advisory services. SERFS may be searched by topic keyword or full–text. There is also a capability to directly link between services and associated data set descriptions to utilize these services. The IDN's OGC portals provide links to OGC related services. The IDN offers an authoring tool to describe services.

2. ECHO Services

ECHO implements a "Service Oriented Architecture" (representing resources as services) for handling services. ECHO services may be published, found and bound by means of an ECHO enterprise–level middleware providing the core services (e.g. Registration, Security, Ordering, Query). ECHO also publishes "extended" services (i.e. which are not ECHO services). ECHO service registries are UDDI based.

It was also demonstrated that existing OPENDAP software could extended to work with ECHO at a limited coast. Mainly, the OPENDAP query service has just to be registered as an ECHO service.

3. CNES Services

CNES is extending its metadata clearinghouse to services. Services are described and organized according to the ISO 19119 Services standard for service metadata and services taxonomy. A service activation function is provided.

4. Product Vitalization through Web Service Chaining

NASA (through the Center for Spatial Information Science and Systems at George Mason University) is experimenting an abstract model designer. It is a Java Applet which enables domain experts to use data types, service types and existing abstract models as basic components to construct abstract models, which represent the domain knowledge of the experts, in a clicking and dragging manner. The experts can register those models into the Catalogue Service for Web (CSW), which will enable the abstract models available for later use. An abstract model can be instantiated and transformed into a concrete BPEL process, which can be executed in a BPEL engine.

5. Mass-market GEO

Google Earth has become the de facto geobrowser for the Web with KML ("keyhole Markup Language) as the language for modelling geographic features. There were 100,000,000 downloads of Google Earth, 30,000 users of the API. 1/3 of world's population can now see their neighbourhood.

Web 2.0 (the "read–write" Web or participatory Web) will allow anyone to publish geodata from many sources, using simple geotagging.

Open Source GIS is reaching maturity. Core geospatial standards available through Open Source (Mapserver, Geoserver, PostGIS,...)

geoRSS is the de facto standard to encode locations in RSS fields.

6. Services for Agriculture Applications

JAXA has shown how the usage of OGC WMS applications can be for benefit of their final users. However, close interaction with the user agencies is required to produce high level information.

Grid Task Team

1. Wide Area Grid Concept

It is felt that the Wide Area Grid concept could be a way to interconnect existing and future grid based systems characterized by different protocols and/or different platforms. However, this concept still needs clarification, especially with regards to large international grid initiatives like the EGEE initiative in Europe. NRSCC, who suggests that GRID should be the acronym of "Glue all Resources into a cyber Infrastructure for Distributed environment", recommends a pragmatic approach for interconnecting heterogeneous systems through dedicated adaptors, as long as there is no de facto middleware standard Some experiments on the gluing between NRSCC grid and Purdue University grid infrastructures have been on the road. In the meantime, CNES with partners like NSAU has started to define a Wide Area Grid prototype.

2. A concept of International Virtual Earth Observatory

The CEOS WGISS experiments on WTF and Grid shows it is possible to provide a persistent CEOS–wide distributed infrastructure to provide a core set of capabilities (data, information,

and tools) to multiple international initiatives. It is suggested to start the International Virtual Earth Observatory (IVEO), coordinated by CEOS WGISS.

IVEO would be a persistent infrastructure providing EO data and information (discovery, access, and analysis) through standard interfaces (OGC, ISO), based on Grid and Web service technology.

3. Product Virtualization in a Geospatial Grid

NASA (through the Centre for Spatial Information Science and Systems at George Mason University) is experimenting a geospatial grid, i.e. a grid for the Geospatial discipline. Users of a Geospatial Grid should be able to access geospatial resources using existing geospatial Web clients without knowing a grid is running. These resources do not need to exist and may be created by the system when the user requests them via a predefined workflow. Such a mechanism has been implemented as a Globus virtual organization on top of a Globus grid running CSW, WCS, WMS services.

4. Cyberinfrastructure for Research and Education

The Cyberinfrastructure for Research and Education and its challenges was presented by Purdue University. The presentation highlighted on :

> Infrastructure building (community clusters, cycles harvest across campus, high speed network links, storage capacity, data collections)

> System interoperability (Integrate computing infrastructures, integrate services)

> Enabling multidisciplinary research and education

The TeraGrid and the NanoHub projects were also presented as projects that make use of the Grid infrastructure at Purdue University.

EOGEO workshop

Participation in EOGEO was strong but has been falling. EOGEO is no longer known outside of a very small community.

It is suggested that the EOGEO be replaced by a "Science and Technology" conference at CEOS level but still with a WGISS focus. This conference would be organized by CEOS member agencies. The first conference could take place in China in the Spring 2008.