COMMITTEE ON EARTH OBSERVATION SATELLITES

Committee on Earth Observation Satellites

Working Group on Information Systems and Services

Interoperability Hanbook

September 2006

Table of Contents

Glossary

Acronyms

CCSDS	Consultative Committee for Space Data Systems
CEOS	Committee on Earth Observation Satellites
CSDGM	FGDC Content Standard for Digital Geographic Metadata
DEM	Digital Elevation Model
DIF	CEOS Directory Interchange Format
DTED	Digital Terrain Elevation Data
EOS	NASA's Earth Observing System
FGDC	Federal Geographic Data Committee
geoTIFF	TIFF format for georeferenced data
GIF	Graphics Interchange Format
GML	Geography Markup Language
HDF	Hierarchical Data Format
IDN	CEOS International Directory Network
ISO	International Organization for Standardization
OAI	Open Archives Initiative
OGC	Open Geospatial Consortium
PNG	Portable Network Graphics
SVG	Scalable Vector Graphics
TIFF	Tagged Image File Format
WCS	Web Coverage Service
WFS	Web Feature Service
WGISS	CEOS Working Group on Information Systems and Services
WMS	Web Map Service
XML	eXtensible Markup Language

Definitions

To be provided later

Foreword

The need for cooperation among the space agencies providing space borne data and the users of these data has been increasing over the past twenty-five years as the result of two factors:

- the development and exploitation costs of space systems;
- the conviction that planet Earth with all its complexity is to be considered as a World natural heritage.

The Committee on Earth Observation Satellites (CEOS) was founded in 1984 in order to bring up this needed cooperation. And, sometime later, within CEOS, the CEOS Working Group on Information Systems and Services was created to address the more specific problem of sharing space borne data and information.

About CEOS

The Committee on Earth Observation Satellites (CEOS) is an international coordinating mechanism charged with coordinating international civil spaceborne missions designed to observe and study planet Earth.

CEOS is recognized as the major international forum for the coordination of Earth observation satellite programs and for interaction of these programs with users of satellite data worldwide.

The three primary objectives of CEOS are as follows:

- to optimize benefits of spaceborne Earth observations through cooperation of its participants in mission planning and in development of compatible data products, formats, services, applications, and policies;
- to serve as a focal point for international coordination of space-related Earth observation activities; and
- to exchange policy and technical information to encourage complementarity and compatibility of observation and data exchange systems.

Fore more information about CEOS, see: http://www.ceos.org.

About WGISS

The CEOS Working Group on Information Systems and Services (WGISS) is one of three subgroups supporting the Committee on Earth Observation Satellites.

WGISS promotes collaboration in the development of systems and services that manage and supply EO data to users world-wide.

The present *Interoperability Handbook* is produced and published by WGISS.

Fore more information about WGISS, see: http://wgiss.ceos.org.

Purpose

Work accomplished by the CEOS Working Group on Information Systems and Services (which replaced in 1994 the former CEOS Working Group on Data) is aimed at providing technical responses to the high level requirements expressed under the CEOS resolution and principles given hereafter.

Within WGISS, agencies join their efforts, identifying the concepts and the underlying technology by which these high level requirements could be satisfied.

This document will go more deeply into these concepts, subsumed under a single one: *interoperability*. Interoperability was once defined by CEOS as *the capability of the user interface and administrative software of one instance of a service to interact with other instances of same type of services*[1].

Services are said to be interoperable if they allow for interoperability as previously defined. And systems are said to be interoperable if they are effective implementations of interoperable services.

This document is the CEOS Interoperability Handbook.

It provides recommendations for the development of interoperable systems, drawn from WGISS 10 year experience. It is a handbook, not an academic essay devoted to the theory of interoperability. It is for immediate use by anyone willing to implement interoperable services in a way that preserves their interoperability.

Developing interoperable systems as implementations of interoperable services makes sense only if there is an existing or expected partnership between organisations developing and/or operating these interoperable systems. By definition, interoperable systems do not behave as if they were insulated. They become *de facto* members of systems of systems linking each of such systems to at least one of its *alter ego*. It results that the CEOS Interoperability Handbook is also for communities who may adopt it as guidelines for systems they want to be (or become) interoperable with other ones. The primary community for this handbook, obviously, is the CEOS community. But others may find benefit in applying it.

CEOS Resolution on Satellite Data Exchange Principles in Support of Global Change Research

CEOS members endorse the following principles relating to satellite data exchange in support of global change/climate and environmental research and monitoring and agree to work toward implementing them to the fullest extent possible. Principles for data exchange in support of other data uses beyond global change/climate and environmental research and monitoring will be developed for CEOS endorsement as a next step.

1. Strictly speaking, this definition was given by CEOS for catalogue services, which are a particular type of services.

1. Preservation of all data needed for long-term global change/climate and environmental research and monitoring is required.
2. Data archives should include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data.
3. International standards – including those generated by the CEOS Working Group on Data – should be used to the greatest extent possible for recording/storage media and for processing and communication of data sets.
4. Maximizing the use of satellite data is a fundamental objective. An exchange/sharing mechanism among CEOS Members is an essential first step to maximize use.
5. Nondiscriminatory access to satellite data by non-CEOS Members for global change/climate and environmental research and monitoring is essential. This should be achieved within the framework of the exchange and sharing mechanisms set up by CEOS Members.
6. Programs should have no exclusive period of data use. Where the need to provide validated data is recognized, any initial period of exclusive data use should be limited and explicitly defined. The goal should be release of data in some preliminary form within three months after the start of routine data acquisition.
7. Criteria and priorities for data acquisition, archiving, and purging should be harmonized.

CEOS Principles on Data Provision

CEOS members endorse the following principles relating to data provision in support of operational environmental use for the public benefit and agree to work toward implementing them to the fullest extent possible within available resources.

1. Criteria and priorities for data acquisition, processing, distribution, preservation, archiving, and purging should be harmonized to take into account the needs of users of data for operational environmental use for the public benefit.
2. Real-time and/or archived data for operational environmental use for the public benefit should be made available on time scales compatible with user requirements and within agency capabilities.
3. CEOS data suppliers should provide (e.g., through the CEOS International Directory Network) easily accessible information about the data and related mission parameters, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data.
4. Recognized standards, to be defined and developed in common, including those generated by CEOS Working Groups, should be used to the greatest extent practical for recording/storage media and for processing and communication of data sets.

5. To optimize the use of data for operational environmental use for the public benefit, CEOS Members should establish appropriate data provision mechanisms.
6. Programs should have no exclusive period of data use except where there is a need to provide for data validation. An initial period of exclusive data use should be limited and explicitly defined. The goal should be release of data in some preliminary form within three months after the start of routine data acquisition.

Abstract

While a stand-alone data system is usually built for a community with specific needs, an interoperable system is built as part of a broader system for different communities sharing similar needs. Interoperable systems are actually systems of systems, with each of their components providing at least one part of the full answer expected by their users.

An interoperable system cannot be implemented in the same way as a stand-alone system because it must be built upon the same underlying concepts as are the other interoperable systems with which it will interact. Thus, implementing an interoperable system requires to follow particular guidelines.

These guidelines applied to Earth observation systems are shown in this handbook as recommendations.

The handbook contains an introduction and several chapters.

The introduction is actually an overview of two basic concepts: the concept of data and information, and the concept of interoperability.

Chapter 1 – Data Preservation – deals with the question of keeping data and information in a state that permits to use them even many years after they had been created. This question must be brought up because interoperable systems are always long living systems.

Chapter 2 – Data Access – deals with the question of discovering, locating and accessing data which may be geographically distributed over many places within interoperable systems.

Chapter 3 – From Data to Information – deals with the question of deriving information – which is what the user actually wants – from the data that are accessible through an interoperable system. The process of deriving information from data is conducted by services. Chapter 3 presents the CEOS view on services applicable to Earth observation data.

Chapter 4 – Infrastructures enabling Interoperability – deals with the question of the underlying infrastructures that permit systems built upon them to become interoperable. It presents the infrastructures recommended by CEOS for the implementation of interoperable systems.

Chapter 5 – Other organizations –deals with the question of interoperability as it is handled by other organizations including CCSDS, OGC, ISO and space agencies.

Introduction

WGISS has been experimenting interoperability over more than 10 years. Lessons learnt from these experiments are gathered in the following chapters and expressed in the form of recommendations.

To make these recommendations more accessible, two fundamental concepts need first to be introduced and clarified:

- the concept of *Earth Observation data and information*;
- the concept of *interoperability.*

About Earth Observation Data and Information

Data is originally the plural form of the Latin word *datum* which means "something given"[2]. Thus, data are "things that have been given". Though data is strictly speaking a word in the plural form, it is accepted in Modern English to use it also as a word in the singular form.

There is a link between data and information. For instance, clouds in the sky are data from which one may derive an information like "rain is coming". Information is actually the result of a mental process on data as "things that have been given". This mental process can also be said as the *process of becoming informed*[3].

There are many instances of data. As such, an object in a museum may be seen as data since it imparts some information. However, the only data that are considered in this handbook are data given as encoded expressions of natural or anthropic phenomena related to planet Earth observed through instruments, many of them being parts of satellite payloads.

These data are called *Earth Observation data* (in short: *EO data*). In general they are given as digitally encoded expressions because they can be submitted to computers only under this form. But they may also be given as traditional documents like printed books or photographs.

Data may be seen as expressions of pieces of a phenomenon. These pieces are selected because they provide a simplified but appropriate view of the phenomenon in a given context, through a modeling process. For instance, the wind over some area may be modelled as a field of vectors giving its local strengths and directions.

A simplified view of a phenomenon is sometimes called a *feature*. Categories of features, called *feature types* may be identified. For instance, the Golden Gate

2. The word *date* has a similar origin: it comes from Medieval Latin *data littera*, first words of the expression used in the past for giving the date when a letter was issued.

3. In some sense, the data "clouds in the sky" contains the information "rain is coming". This explains why, very often, data and information are considered identical. For a thorough analysis of the concept of information, see: Information and Information Systems, Michael Buckland, 1993, ISBN 0-313-27463-0.

bridge in San Francisco is a instance of the feature type *bridge*. Features and feature types are categorized by *attributes*. The name of the bridge, its height, the material used to construct are attributes. When a phenomenon is viewed as the variation of a function within a spatiotemporal domain (for instance the variation of a vector), its view is also called a *coverage*.

One peculiarity of EO data, in comparison with other types of data (like astronomical data), is that they always refer to some position on the Earth. This is due to the fact that a phenomenon appears at some place on the Earth. This place generally has a 2- or 3-dimensional spatial extension. In addition to its spatial domain, a phenomenon can also have a temporal domain given, for example, by a beginning date and an ending date. This makes EO data intrinsically complex data.

Data having similar characteristics (for instance wind vectors as measured by a single instrument) may be grouped. The result of such a grouping is called a *dataset*. On a computer, a dataset is often handled as collection of files. A dataset may be processed or formatted for future use. In this case it is called a *data product*.

An instrument rarely observes directly the phenomenon of interest. It usually only observes some basic phenomenon. The data relevant to the phenomenon observed by the instrument are later transformed by application of an appropriate algorithm into data which are then taken as data relevant to the phenomenon of interest. For instance, data relative to wind vectors over the ocean are derived from the data relative to an another phenomenon, namely the response given by a radiometric signal after reflection on the ocean surface. The response depends on the profile of the ocean waves which, in turn, permits to derive the wind vectors. Several algorithms may be applied in cascade to an initial data product to get the desired final product. It results that algorithms should be attached to the data since they contribute to the description of the phenomenon of interest.

In the past, CEOS has identified five levels of data products. They are recalled hereafter:

- **Raw Data –** Data in their original packets, as received from a satellite.
- **Level 0 –** Reconstructed unprocessed instrument data at full space-time resolution with all available supplemental information to be used in subsequent processing (e.g., ephemeris, health and safety) appended.
- **Level 1 –** Unpacked, reformatted level 0 data, with all supplemental information to be used in subsequent processing appended. Optional radiometric and geometric correction applied to produce parameters in

physical units. Data generally presented as full time/space resolution. A wide variety of sub-level products are possible.

- **Level 2** – Retrieved environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the level 1 source data.
- **Level 3** – Data or retrieved environmental variables which have been spatially and/or temporally resampled (i.e., derived from level 1 or 2 products). Such resampling may include averaging and compositing.
- **Level 4** – Model output or results from analyses of lower level data (i.e., variables that are not directly measured by the instruments, but are derived from these measurements).

Each level represents a step in the abstraction process by which data relevant to physical information (raw, level 0, level 1) are turned into data relevant to geo-physical information (level 2, level 3), and finally turned into data relevant to thematic information (level 4).

About interoperable services and systems

The process of transforming data may be provided as a *service*. The ISO 19119 standard[4] defines a service as the distinct part of a functionality that is provided by an entity through *interfaces*. The same standard defines an interface as a named set of *operations*, and defines an operation as the specification of a transformation. An operation has a name and a list of input and output parameters. A service may consist of one single operation. It may also consist of the chaining of several operations.

The above definitions do not put any restriction on the way operations are implemented. Operations can be – and often are – manual operations. However, in this handbook we are considering only operations that are implemented on computers.

Different entities may implement in different ways the same service. In this case, each entity makes the service available through one of its own set of interfaces.

Two or more implementations of the same service are interoperable if the invocation of one implementation entails the invocation of the other implementations. The values of the output parameters of the last operations in each service chain are merged and make the complete output of the service that was first invoked.

Interoperable systems are systems that implement interoperable services.

4. ISO 19119:2005 – Geographic information – Services

To clarify the above statements, let us consider the following small catalogue service with one single operation:

operation name	get_metadata
input parameter name	time_period (mandatory)
input parameter name	keyword (optional)
input parameter name	metadata_description_type (mandatory)

The catalogue service operation retrieves descriptions of existing EO data products (these descriptions are also called *metadata*).

Descriptions are retrieved only if the corresponding data products have a temporal domain that intersects the time period given by the *time_period* parameter. A time period may be expressed, for instance, by the indication of a beginning date and of an ending date in the Gregorian calendar. Additionally, if the corresponding data products may be characterized by a keyword, the only descriptions that are retrieved are those whose corresponding data products match the keyword given with the *keyword* parameter. This constraint applies only if there is a *keyword* parameter. And finally, among the various descriptions that are retrieved, only the description with same type as the type given with the *description_type* parameter are definitively retrieved. Common description types are *brief*, for short descriptions, and *full* for comprehensive descriptions.

Such a catalogue service may be implemented in many ways. For instance data product descriptions may be stored in a relational database, retrieved using SQL requests built from the get_metadata operation input parameters and sent to the service user as a text file.

If such an implementation is to become interoperable with another one, several conditions must be taken into account:

- since the request by which one service is invoked must be forwarded to the other one, both implementation sites must be connected in some way;
- the incoming service request must be encoded by the first service implementation according to an encoding mechanism recognized by the second implementation;
- conversely, the descriptions retrieved by the second service implementation must also be encoded according to an encoding mechanism recognized by the first implementation.

For instance, if both services are implemented as http servers, and assuming there is an internet connection between the server sites, the http protocol structures could be used for the transportation of requests/replies between them. Both implementers will need to agree on common encoding schemes for the messages that will flow from one implementation to the other. They may select some encoding schemes amongst well established standards. For instance, dates may be encoded according to the CCSDS recommendations for date/time encoding, and metadata that are retrieved by the servers may be encoded according to the CEOS Directory Interchange Format (DIF).

Beyond the technical solutions which may be elected to achieve interoperability between the two service implementations, the point to emphasize is that interoperability requires the definition of additional *private* service interfaces by which systems implementing services privately provide interoperability services to each other. These private services (in our example, the ability for one service implementation to forward get_metadata requests to the other one) extend the public service interfaces by which the implemented service is publicly seen by the ordinary user. This is one reason why making an already existing service implementation fully interoperable with other ones is always a challenge.

The first of the above conditions deals with interconnection.

The second and third conditions deal with structure and encoding, i.e. with *syntax*.

There is a fourth condition, dealing with *semantics*: implementers need to share a common understanding of the service they want to implement.

For instance, implementers of the above catalogue service would need to agree on a common definition of data product descriptions, both in their brief form and in their full form. Such a definition will become a *de facto* standard for the two implementers. They will also need to agree on the handling of the optional *keyword* parameter value: should it be checked against a list of valid keywords or just handled as free text? One implementer may also not completely implement the service. For instance, he may not implement the optional *keyword* parameter part of the catalogue service because the data product descriptions on his side all belong to a single category and discrimination by keyword does not make sense. He must however be prepared to semantically understand and syntactically recognize the *keyword* parameter in an incoming *get_metadata* request, in order to ignore it without rejecting the request.

And there is finally a fifth condition: *time*.

Most interoperable services are designed as long term responses to fundamental needs common to many user communities. Discovering data, locating data, accessing data, using data are examples of such fundamental needs of EO data users. Thus, most interoperable services are long living services, often defined by standards, and all objects handled through interoperable services must be set up with a perspective of long term use and reuse.

In some sense, interoperability should be considered as goal to reach rather than a property a system may exhibit or not. In the past, CEOS has defined levels of interoperability. They are recalled hereafter:

- **Level 1.** A user accessing one service is routed directly to another related service. For example a user identifying a relevant data set in a directory is routed directly to the associated inventory service, or a user of one inventory service is routed automatically to another related inventory to extend a search. At this level, the user interface differences between the routed services are not hidden from the user.
- **Level 2.** This is similar to level 1 except that a user accessing one service is routed directly to another related service, and context information is passed at the same time. The context information typically includes information concerning the user, their interests, and the activity on the

original service (for example a search query) which can be interpreted by the new service to assist the user in making further queries. For example search query criteria entered at one site can be passed to another, translated and loaded as a search query for the user to activate. At this level the user is still exposed to the differences in operation between the two services, but related information can be shared.

- **Level 3.** At level 3, the user is hidden from the specific operations of each service. A service request is routed automatically to one or more services, and the results from each returned to the original access point. At this level (and levels 1 and 2) each service is likely to use a different data model and the query may need translation to reflect these differences. Similarly, the query results may also reflect the individual service data models and may or may not be translated before they are presented to the user. At this level the user is no longer exposed to the differences between service operation.
- **Level 4.** At the highest level of interoperability a single data model is assumed to apply to all services within an interoperable system. Thus a query entered by a user can be executed using a single distributed database. This level of interoperability permits direct database operations such as union and joins between metadata held by different catalogue services and for operations. The interoperable system is thus seen as a single database system by the user.

Several interoperability levels may coexist in a system made of interoperable systems:

- Some systems could just provide level 1 interoperability: the user of one system may be linked by this system to another system via a url address but he will have to rebuild himself his request within the second system interface. This is the weakest level of interoperability but it is often the easiest way to make existing systems interoperable with other ones.
- The small interoperable catalogue service system example would provide interoperability at level 3, since the get_metadata request is automatically routed to the other site and the results sent back to the user.
- Some organizations could locally agree on a common data model for a given discipline and offer level 4 interoperability for data relevant to this discipline through a distributed database.

Data Preservation 1

1.1 Introduction to Data Preservation

Data preservation is composed of multiple elements like data management, catalogue management ,and data providing.

After creating products by processing L0 data, it will archive to storage system, at the same time data information called metadata will be created and cataloged. Catalogued metadata requires one-to-one correspondence to data products.

When requesting products, user will find data by searching catalogue then access to data via network.

Each catalogue and data nodes are required open standard interface.

Data is required to keep for the long term.

1.2 Data Archiving

Data archiving is to save processed data to the storage. It requires some of elements to use a ”real data archive”. It means form the user point of view, ”archiving data” always available to search, access and easy to use its system. User can access multiple data nodes as if accessing one node.

To keep the data for the long term, virtual storage system will required on each node.

So Data archiving requires following conditions.

- Archiving products to storage and registering the metadata to the catalogue.
- There is a link to data product in the Metadata and metadata format requires ISO standard format (ISO 19115).
- Access protocol also requires standard (z39.50), it actualize as if using one node in spite of accessing multiple node
- User can access to data products via network by searching catalogue.
- For the long term archive, a data not frequently accessed will move to media in the robotics tape machine, a data frequent accessed will on the disk system, virtual storage system will combines both raid disk and robotics tape machine.

1.3 Data Formats

Ideally user can select data format among the major formats.

Data Access

2

2.1 Introduction to Data Access

A user may want to access data, i.e. to retrieve data from the various repositories where they are stored. This implies that the user knows how to identify the data, locate the corresponding data repositories, and interact with these repositories to retrieve the data.

Generally, the user does not know directly the identity of the data but he is able to provide some characteristics of the data. The identity of the data may be obtained if there exist a list showing together existing data identities and their corresponding data descriptions to which the user provided characteristics may be compared. Such a list is called a catalogue[5]. A catalogue may be as simple as a printed document. It may also be organized as a complex computer system with sophisticated search and retrieve functionalities[6].

While the characteristics by which the user knows the data may be more or less precise, the descriptions that are found in the catalogue must be as precise and exhaustive as possible (the catalogue is actually a registry containing official records) and identify without ambiguity the data of interest to the user. In particular, the data description from the catalogue must contain the information needed by the user to request and get the data. Semantically, this information is the path which leads to the record(s) containing the data. It can be a simple URL in which case there is no complicated interaction (at least from the user point of view) with the repository hosting the data. It can also be a complex data order and data retrieval specification (sometimes including the payment of fees). In this case, the interaction with the repository may be complicated and need several more or less automated steps: contact the data custodian, order the data, pay the fees, retrieve the data with the appropriate protocol.

In the following paragraphs, we discuss in more details the three basic steps that are needed for data access:

- describing data
- searching data
- retrieving data

5. According to the American Heritage Dictionary, a catalogue is *a list or itemized display, as of titles, course offerings, or articles for exhibition or sale, usually including descriptive information or illustrations*.

6. A catalogue system may comprise several single catalogues interconnected via an interoperability protocol. An example of such a catalogue system is the eoPortal Catalogue (http://catalogues.eoPortal.or g) whose catalogues are interconnected via the CEOS Catalogue Interoperability Protocol (CIP).

2.2 Describing Data

About Data Description

Describing Data is providing information about data.

Examples of information about data are:

- the location of the data repositories
- the title by which the data is known
- the data content and the data purpose
- the data spatial and temporal domains

The technical word for data descriptions is *metadata*.

Describing data through metadata consists in providing views on the data. There may be many different views on data. Some of them may be very high level views, just providing information about the existence of data through a title and a short abstract. Others may go into details like, for instance, the various ways the data may be retrieved from each of its repositories.

Metadata provide a mediation between a user searching information and the repositories where the data containing the searched information may be found. In this sense, metadata act as data surrogates: while handling their metadata, the potential user can check if the data would satisfy his needs without having to physically extract them from their repositories.

This has some implications on the way metadata must be constructed:

- metadata must describe data in a way understandable by the metadata user; the words that are used in metadata must be part of the vocabulary expected by their users (vegetation specialists do not refer to the same concepts – and thus do not use the same words – as ocean specialists).
- there may be several metadata for the same data, depending on the metadata target audience; metadata for laymen are different from metadata for experts; if the same data are of interest in several disciplines, there may be several metadata for the same data, each one constructed according to the peculiarities of the related discipline.
- conversely, data that differ only at some level of details may be organized as a single collection; description would apply at collection level, not at data level.

In most cases, metadata may be seen as a document linked to the data (or data collection) they describe. Such a document consists of chapters and of paragraphs, each one devoted to a particular view on the data. It mainly contains printable text but may also contain (or be linked to) graphical representations like images. Metadata that are digitally encoded may be electronically exchanged. XML has become a very common technology for the digital encoding of metadata.

A community may agree on a metadata template. Such a template indicates the way metadata should be constructed by data providers and data custodians for the members of the community, i.e. the template gives the names and the topics of the chapters and paragraphs the metadata should or may contain. A large community may define a very general metadata template. This template may be further reused and profiled for their specific needs of smaller communities that are part of the larger one.

Metadata templates may become international *de jure* or *de facto* standards and there may also be metadata encoding standards.

There should always be a link from the metadata to the data they describe and, conversely, there should always be a link from the data to their metadata.

Metadata may be recorded sequentially in a single file which can be printed or displayed. Metadata may also be gathered in a database (usually, a community that owns hundreds or more metadata makes them available only through a database against which the user applies queries).

CEOS recommended metadata standards

Recommendation 1

The CEOS Directory Interchange Format (DIF) should be used for the description of science Earth Observation data.

Rationale

The DIF, which was created through international consensus has become the de facto standard for describing Earth Observation datasets across many science disciplines. The DIF not only standardizes the metadata structure but also the metadata content: for example, the specification of personnel or the specification of dates. It is developed and approved through the CEOS Interoperability Forum.

Robust software is available for writing DIF compliant metadata. DIF metadata are now encoded as XML texts and may easily translate to other metadata standards like Dublin Core, ISO 19115, CSDGM.

Recommendation 2

The CEOS International Directory Network (IDN) should be used to host DIF compliant metadata.

Rationale

DIF compliant metadata may be encapsulated into the CEOS International Directory Network (IDN). The IDN is an operational database for DIF compliant metadata with powerful search and retrieve capabilities. It currently holds more than 16,000 metadata.

Note

Limited views to the IDN metadata content, for instance, a view to an organization's metadata, are offered through dedicated IDN portals. Hence an organization can easily build a dedicated catalogue containing DIF compliant metadata without having to start from scratch.

Recommendation 3

Description of geographic data should be done in conformity to the ISO 19115 metadata standard.

Rationale

Whilst there is a wide overlap between them, Earth observation data and geographic data should be considered distinct. Geography is the *study of the earth and its features and of the distribution of life on the earth, including human life and the effects of human activity*[7]. Thus, geographic information may be very complex, sometimes much more complex than ordinary scientific information.

It results that the descriptive power of a metadata format like the DIF may not be sufficient for geographic data. "ISO 19115 - geographic information -- metadata" is the current international standard for the description of geographic data.

Note

A community may develop an ISO 19115 metadata standard profile. A profile is an extension of the standard[8]. Conformity to a registered profile entails conformity to the ISO 19115 metadata standard.

2.3 Searching Data

About Data Search

We assume that the user who wants to retrieve data at least knows some characteristics of the data. These characteristics are generally based on the user's knowledge and expectations. For instance, a user interested in the radiation flux received or emitted by the Earth at top of the atmosphere over Europe may characterize the data he wants to retrieve by the words "Earth Radiation Budget" and the words "Europe": "Earth Radiation Budget" is the name of the scientific discipline which studies the radiation received and emitted by the Earth; "Europe" is the name of a geographic region but "Europe" could have also been defined (perhaps with less ambiguity) by the geographic coordinates of a polygon surrounding the geographic region known to be Europe. Hence, the data characteristics provided by the user is made of a keyword ("Earth Radiation Budget") and a spatial extent (given either by the geographic name "Europe" or by a sequence of geographic coordinates). Both the keyword and the spatial extent characterize the data. Hence, keyword and spatial extent are implicitly linked by the Boolean operator AND. If the user is only interested in short wave radiation, he may want to add the keyword "Short Wave Radiation" and link it to the previous ones with the Boolean operator AND.

The user must then find a catalogue that understands these data characteristics. Obviously, if no catalogue knows the keyword "Earth Radiation Budget", the user will never find what he wants, even if the catalogue contains very detailed

7. Definition given by the *American Heritage Dictionary.*
8. The standard gives rules for creating a profile.

descriptions of data relevant to the Earth radiation budget but unfortunately only known by the catalogue only under the keyword “Atmospheric Radiation”. The consequence is that a catalogue should always be queried with the keywords (or, in a broader sense, with the vocabulary) the user is used to use within his community.

The technical name of the index of all terms in use within a community is “thesaurus”. A thesaurus is a structured index, showing all the relations between terms (e.g.: a term may be the synonym of another term, a term may a specialization of another one, etc.), thus suggesting the possible navigation schemas through the terms[9]. A thesaurus containing geographic terms (names of countries, cities, etc.) is sometimes also known as a “gazetteer”.

CEOS Recommendations for Data Searching

Recommendation 4

A catalogue should always link each of the keywords available for data searching to the appropriate thesaurus.

Rationale

Since they are the most common way to browse through catalogues, keywords must be very strictly controlled in order to avoid as much as possible ambiguity. If there is an ambiguity in the definition of keywords, the user may retrieve unwanted data or, which is worse, not find valuable data.

Note

Keywords from different thesauri may be used in a single search.

Recommendation 5

For science data, catalogues should use to the far extent possible the keywords defined by the CEOS International Directory Network.

Rationale

The keywords defined by the CEOS International Directory Network are well accepted by most of the scientific communities because they have undergone a rigorous selection process under the auspices of top representatives of these communities.

Note

Using the keywords defined by the CEOS International Directory Network does not imply for a catalogue to comply with the CEOS Directory Interchange Format for the metadata.

9. Guidelines for the establishment of monolingual thesauri are given by the international standard ISO 2788. Guidelines for the establishment of multilingual thesauri are given by the international standard ISO 5964.

2.4 Retrieving Data

About Data Retrieval

A catalogue does not necessarily directly give access to the data but at least it gives the guidelines the user should follow to access the data. These guidelines can be found in the metadata[10]. To access the data, the user must go through a special entity: a data distributor.

A data distributor may distribute only one kind of data. Or he may distribute several different kinds of data. There may also be several different data distributors for the same kind of data. Sometimes, several data distributors make a federation of data distributors, in which case the user interacts with only one of them who forwards the requests for data retrieval to the other ones.

Whatever the data distributors' organization, the user will always have to prepare a data order request, submit it to a data distributor, wait for the order processing and retrieve the data.

Order Preparation

The user selects the data from the metadata and prepares an order to be submitted to one of the existing data distributors. The identity of the data distributor can be found in the metadata.

Order preparation may be as simple as just selecting a URL if the data are available on line from a http server. But usually order preparation is much more complicated and includes the following steps:

- identify the data distributor

  There may be several distributors for the same data. The user must select at least one data distributor from the metadata[11]. A data distributor may distribute data on line or off line. Large data may be distributed only off line.

- get the order instructions

  The order instructions should be found in the metadata or, at least, the metadata should give the identity of a resource holding these instructions (the resource may be the web site of the data distributor).

- set up the order according to the order instructions

  The order contains items like the identity of the data and, if there is a choice, the medium that will be used to transfer the data to the user with the appropriate format. The medium may be a physical medium if the data is to be transferred off line (e.g. cdRom, digital linear tape). It may be an electronic link to an ftp server which will host the data. The ftp server may be an ftp server owned by the data distributor (in which case the data distributor must allow the user to be a registered ftp server user) or it may be an ftp server owned by the user (in which case the user must allow the

10. For instance, the ISO 19115 metadata standard has a section entitled "Distribution Information" giving details about the distribution of the data to the user, including the type and format of the media used for the transfer of the data to the user.

11. As mentioned earlier, the data catalogue may sometimes act as a data distributor.

data distributor to be a registered ftp server user and he must send him the required login information).

The order also contains items like the address where the medium will be delivered or the address where the data distributor will send the invoice – if any – related to the ordering.

Order Submission

Once it is prepared, the order must be submitted to the data distributor in the way indicated by the order instructions, e.g. by mail, by electronic mail, by fax or directly by updating the data distributor's web site. User authentication may be requested by the data distributor and the order may be enciphered during the transfer to the data distributor. The order is evaluated by the data distributor who accepts the order unless he cannot process it for some reason, in which case the order is rejected. The user should be informed of the acceptance or rejection of the order, in a way indicated in the order instructions (sending an Email to the user is a common way). The order is given an identifier by the data distributor for further order retrieval by the user.

Order Processing

Once it is accepted, the order is processed by the data distributor. Order processing may be as simple as sending to the user a web page containing the data that were ordered. However, order processing by the data distributor is generally much more complicated and includes following steps:

- find in the data repositories the various items that make the data ordered by the user; these items are known as granules, i.e. the smallest parts of the data retrievable from a repository, and are often packaged as files.
- extract the granules from the repositories and create the dataset expected by the user, i.e. the collection of all the data ordered by the user. This collection must also contain the data descriptions that are needed for the utilization of the data. These data descriptions provide the syntactic view of the data (whilst the metadata provide the semantic view). They actually are the *instructions for use* of the data. If they are not available in the dataset, the user cannot use the data.
- transfer the dataset to the user as indicated by the order.

Additionally, the data distributor maintains a record of all orders received from the user with their status. The user may access this record, check the current order status. He may also cancel an order being processed. An order may have one of the following status:

- accepted – the order is accepted by the data distributor
- rejected – the order is rejected by the data distributor
- being processed – the order is being processed by the data distributor
- waiting – the order is being processed but cannot be completed unless some event has occurred[12]
- suspended – the order is suspended because something prevents its completion[13]
- cancelled – the order has been cancelled (either by the user or by the data distributor)

Order completion

The user is informed by the data distributor that the order has been completed. Data should be made available to the user immediately after transfer.

CEOS recommendations for retrieving data

Recommendation 6

The development of a data ordering facility should follow the recommendations of the CEOS Catalogue Interoperability Protocol for data ordering.

Rationale

At the end of the 1990's, CEOS has experimented the CEOS Catalogue Interoperability Protocol (CIP). The CIP was the result of an initiative to develop a protocol so that a number of international agencies could make their Earth Observation data, and related data, available in coherent manner to their users communities[14].

CIP documentation contains a very detailed and exhaustive description of the full ordering process from all the points of view that are needed for the implementation of a data ordering system: user requirements, system specification, system design.

Recommendation 7

Data that may be accessed online should at least be accessible via the ftp protocol.

Rationale

The ftp protocol is widely used in the world. It should be available by default for online data retrieval.

12. A data distributor may offer a subscription service, i.e. an order submitted by a user may be related to data that do not already exist but are known to be available in the future. In this case, the order cannot be processed immediately.

13. If the data must be transferred to an ftp server and there is not enough space on this server, the order processing must be suspended.

14. Source: CEOS Catalogue Interoperability Protocol – (CIP), Specification – Release B

Recommendation 8

The data descriptions provided with the dataset should follow the recommendations of the ISO 15889:2003 standard (Space data and information transfer systems -- Data description language -- EAST specification)

Rationale

EAST is a powerful data description language defined by space agencies within the Consultative Committee for Space Data Systems (CCSDS). It allows to describe data at a very detailed level (including endianness). Tools exist to create or interpret datasets from their EAST descriptors and generate automatically the related documentation.

From Data to Information 3

3.1 Introduction to Interoperable Data processing

Earth Observation data are often required to be served in the form of information for end users. Interoperable technologies will enhance the capabilities to deliver information to the end users generally through application systems which interact with dispersed systems to perform services.

To provide information derived from EO data through the interoperable systems, three mechanism will support the process.

- The first one is data serving mechanism to provide data contained in the service system to other systems.
- The second one is data processing mechanism to produce higher processed data such as value added data. Data processing function to create higher level data could be embedded as an internal function within a data providing system or a client system, however another method is to use a data processing system which works interactively with multiple dispersed data providing systems and dispersed client systems.
- The third one is a mechanism to find and bind unknown multiple services to interact with. Service registry is the place to register services to be found and bound by other systems. Through a registry, sharing broaden services could be achieved by connecting them dynamically. As a consequent, registries would enhance opportunities to share services to create information and disseminate it.

3.2 Interoperable Data Providing via Internet

Outline

Specifications in the geospatial domain, such as those developed by OGC and the ISO 19100 series are also applicable to Earth observation data. These standards allow EO data and information to be retrieved in a interoperable manner among dispersed service systems on the Internet.

Among OGC standards, implementation specifications for data providing purposes are specified in WMS (Web Map Service), WFS (Web Feature Service), and WCS (Web Coverage Service) specifications. Those specifications define interfaces to provide data, low-level processed EO data and highly processed data which are turned into thematic information such as thematic image derived from EO satellite data.

A data providing system provides data held within the system, according to a request received from a client system, by transforming the data into a requested geospatial configuration through on-the-fly reprojection, resampling, and subsetting. The results are ingested into client systems for general purposes such as to display maps in an interactive web mapping system, or for further processing in a dynamic data processing system.

Technology Overview

A service, in this case a WMS or WFS or WCS data providing service, consists of multiple operations. To activate a service, operations are invoked from a client system and received, parsed and performed by the service system to return a result to the client. The main operations of the WMS, WFS and WCS are the *GetCapabilities* operation and the operations to provide data such as the *GetMap* operation in the WMS.

A *GetCapabilities* operation provides information in a form of XML document about content data and acceptable variables of given parameters to formulate valid requests in the next operation. This service-level information called a service metadata.

Then, by using parameters shown in the service metadata, through data acquiring operations such as the *GetMap* operation in the WMS a client system can request data of interest in a specified projection and a specified format.

Example of parameters used in the *GetMap* operation are as followings:

- server url,
- layername (data name),
- supported format,
- supported coordination systems
- supported geographical range (boundary box)
- supported time range
- background transparency

Usually, application systems will perform those operations implicitly from the end users but will provide user interface to help end users to request data on demand.

The main character of WMS, WFS, WCS in OGC specifications are summarized as below.

- WMS (Web Map Service) produces maps of georeferenced data, and allows a client system to get map images from multiple Web Map Service systems on the Internet. “map” is defined as visual representation of geospatial data. Maps are generally rendered in a pictorial format such as PNG, GIF, and so on, or occasionally as vector-based graphical elements such as SVG.
- WFS (Web Feature Service) allows a client system to retrieve geospatial data encoded in GML (Geography Markup Language) from multiple Web Feature Service systems on the Internet in a similar fashion with WMS.

WFS also supports operations such as *INSERT*, *DELETE* on geographic features.

- WCS (Web Coverage Service) allows a client system to retrieve geospatial data as "coverages". A WCS system provides coverage data in forms that are useful for client-side rendering, multi-valued coverages, and input into scientific models. Coverages will be encoded in digital geospatial data format such as GML, GeoTIFF, DTED, HDF-EOS, and so on, to represent spatiotemporal varying phenomena.

(details about OGC WMS, WFS, WCS spec can be found in http://www.opengeospatial.org/specs/?page=specs)

a) Recommendation

(Recommendations need to be examined.)

Recommendation 9

In the field of Earth Observation, WMS, WFS and WCS are recommended to provide spatiotemporally subsetted and projected data in a web-based dispersed environment in an on request basis, for purposes such as web mapping and dynamic data processing.

WMS is to be recommended to provide image data in popular image format, or vector-based graphical format such as SVG, WFS for feature data such as simple vector data in GML, WCS are to deliver low-level data, DEM data, and other forms of data which can be ingested into further processes to be performed in a client system, or in other service systems, to render the acquired data to visualize or to analyze spatiotemporal phenomena.

Rationale

(i) These standardized services make the system invokable from other systems and let the system provide data in an on request, on-the-fly basis.

(ii) By adopting these standards as a basis for a data providing service, benefits like the following are obtained:

- become free from software/hardware issues between a client system and a service system.
- compatibility will be increased among similar systems, provided by different service entities.

b) Known constraints

(Below is an example)

From WGISS experiences, several limitations have been recognized to utilize WMS, WFS, WCS services.

(i) Differences in implementation

Although interfaces between a client system and a service system are defined by standards, ways to implement operations may varies among different systems. For example, to provide time series data, a system may use *TIME* parameter to specify time condition, while other system may determine time dimensional information within the *LAYERNAME* parameter by combining layer name with time stamp, sometimes of user defined time stamp. Therefore, it is possible that implementation may vary among systems and service metadata should be examined first to connect systems.

3.3 Interoperable Data Processing via Internet

(T.B.D)

3.4 Access to Multiple web-based Interoperable Service Systems

(T.B.D)

Infrastructures Enabling Interoperability

4

4.1 About Infrastructures enabling Interoperability

To be written by all task teams

4.2 Technology enabling Interoperability

To be written by all task teams. This paragraph should discuss technology issues like Z39.50, http, web services, grids.

4.3 CEOS Infrastructures enabling Interoperability

To be written by all task teams

The Interop Forum

The Interoperability Discussion List (Interop Forum) is designed for the international community to discuss issues that relate to the DIF and to the SERF and their retrievals through the IDN.

The IDN enables interoperability through compliance to standards like the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (PMH) or international character sets, and to guidelines like the American Disabilities Act (section 508).

The IDN offers discussion lists through the CEOS plone web site and shares information through the CEOS IDN Newsletter. It is also engaged in many partnerships and collaborations.

The eoPortal

Others

Other Organizations 5

This chapter will contain information about interoperability issues as they are handled by other organizations, like CCSDS, OGC, ISO and agencies.

CEOS Recommendations for Interoperability

1	The CEOS Directory Interchange Format (DIF) should be used for the description of science Earth Observation data.
2	The CEOS International Directory Network (IDN) should be used to host DIF compliant metadata.
3	Description of geographic data should be done in conformity to the ISO 19115 metadata standard.
4	A catalogue should always link each of the keywords available for data searching to the appropriate thesaurus.
5	For science data, catalogues should use to the far extent possible the keywords defined by the CEOS International Directory Network.
6	The development of a data ordering facility should follow the recommendations of the CEOS Catalogue Interoperability Protocol for data ordering.
7	Data that may be accessed online should at least be accessible via the ftp protocol.
8	The data descriptions provided with the data set should follow the recommendations of the ISO 15889:2003 standard (Space data and information transfer systems -- Data description language -- EAST specification)
9	In the field of Earth Observation, WMS, WFS and WCS are recommended to provide spatiotemporally subsetted and projected data in a web-based dispersed environment in an on request basis, for purposes such as web mapping and dynamic data processing. WMS is to be recommended to provide image data in popular image format, or vector-based graphical format such as SVG, WFS for feature data such as simple vector data in GML, WCS are to deliver low-level data, DEM data, and other forms of data which can be ingested into further processes to be performed in a client system, or in other service systems, to render the acquired data to visualize or to analyze spatiotemporal phenomena.

Index

A

C

D

E

F

G

H

I

M

N

O

P

R

S

T

W

X