Instructions for GEOSS Data Providers
(How to Place Tags in the Metadata for GEOSS Data-CORE)

Point-of-Contact: For assistance with metadata tags for GEOSS Data-CORE, please submit a request for assistance to the Standards and Interoperability Forum (SIF) by filling out a short form at http://bit.ly/MS8COZ

The GEOSS Data Collection of Open Resources for Everyone (GEOSS Data-CORE) is a dynamic collection of data that has been pledged by GEO Members and Participating Organizations, and that is made available to the GEOSS community with full, open and unrestricted access.

The GEOSS Data-CORE is one of the most important added values of GEOSS to the global community to address the priority Societal Benefit Areas identified by GEO. This note provides practical advice on how to turn the pledges made by your organization into discoverable GEOSS Data-CORE data for the user community. Further guidance will be provided as we learn from operational practice.

The official definition of GEOSS Data-CORE is:

> *"The GEOSS Data Collection of Open Resources for Everyone (GEOSS Data-CORE) is a distributed pool of documented datasets with full and open unrestricted access at no more than the cost of reproduction and distribution"*

where:

- **"*full access*"** means that all the data in the GEOSS Data-CORE can be accessed, used, and redistributed;

- **"*open access*"** means that data providers may charge at most the cost of reproduction and distribution of the data, although it is expected that in most cases the data in the GEOSS Data-CORE will be made available at no cost;

- **"*unrestricted access*"** means that no restrictions are placed on the access to, or use and redistribution of, the data in the GEOSS Data-CORE. It should be noted that the following two conditions may be placed on data registered in the GEOSS Data-CORE by data providers: (i) attribution and (ii) user registration. These are not considered to represent restrictions on the access to, or use and redistribution of, the data.

There are four pieces of information identifying data as part of the GEOSS Data-CORE. Such information has to be expressed as special character strings (tags) and placed in the metadata (metadata tags). In the following sections we describe the tags, their placement in the metadata, how to register data and data catalogs in the GEOSS Data-CORE, and how to discover the data using the GEOSS Common Infrastructure (GCI).

Metadata Tags

GEOSS Data-CORE metadata tags must be placed in the metadata in order to:

(i) allow users to discover and identify the GEOSS Data-CORE data, and

(ii) inform the data users of any conditions attached to the use of the data.

The four metadata tags that describe the data in the GEOSS Data-CORE are:

1. **geossDataCore** – this tag identifies the data that belong to the GEOSS Data-CORE. Thus, data without this tag will not be identified by data users as GEOSS Data-CORE data.

2. **geossNoMonetaryCharge** – this tag identifies the data that are available at no cost, (i.e., the cost for reproduction and distribution has been waived by the data provider, and the data are made available for free access and use).

3. **geossUserRegistration** – this tag identifies the data requiring data user registration and login in order to gain access to the data. At the moment, data providers requiring user registration and login should use their existing mechanisms to accomplish this task. A central user registration and single sign-on solution for the GEOSS is being developed and tested in 2012.

4. **geossAttribution –** this tag identifies the data requiring or recommending attribution. The data user should make sure that attribution is provided whenever the data is used, redistributed, derived from, etc. Data providers requiring attribution should ensure that the information they want to be used for attribution is included in the metadata. A Citation standard is being developed and tested in 2012.

Metadata Tag Placement

The metadata tags are not case-sensitive. There are a variety of metadata standards used by various data providers and scientific communities. Although placement in any free-text searchable field of the metadata will result in the tags being discovered by the GCI, and made known to the data users, we show specific examples of preferred metadata fields for the following standards (refer to ANNEX I for the technical description of placing the GEOSS Data-CORE tags in the metadata):

- ISO 19139 metadata
- Dublin Core metadata
- OpenSearch
- WCS, WFS, WMS, WPS metadata
- GBIF metadata
- THREDDS metadata
- NetCDF metadata
- DIF metadata

Components and Services Registry (CSR) Registration

The GEOSS Components and Services Registry provides a formal listing and description of Earth observation systems, data sets, models, and other services, tools, and resources that contribute to the GEOSS. The CSR provides a registration process that resource providers to GEOSS can follow to register their resources, including GEOSS Data-CORE data.   In particular, the CSR registration process has a step that allows the GEOSS Data-CORE data provider to specify, using checkboxes, which of the four metadata tags discussed earlier should be associated with the data being registered. This step has the title "Resource Sharing Properties."

If the resource being registered at the CSR is a catalog, then it is important that the catalog be homogeneous with respect to the GEOSS Data-CORE tags. This means that all data resources referenced by records in the catalog must share the same GEOSS Data-CORE tags.   The concept is discussed in detail below in the *Metadata and Catalog(ue)s* section.

The success of the GEOSS Components and Services Registry will depend on the commitment of GEO Members and Participating Organizations to input and update their registration details on a regular basis. Registration involves completing a standardized form on-line, and it only takes a few minutes.

Please use the link below to register your components, services, data, and other resources: http://geossregistries.info/geosspub/

Metadata and Catalog(ue)s

For the purposes of the metadata tags discussed earlier, a homogeneous catalog is a catalog where every metadata record in the catalog is associated with the same metadata tags. For example, if all records in a catalog are meant to reflect the geossDataCore tag and none of the other metadata tags discussed, then the catalog is considered homogeneous, and can be registered as GEOSS Data-CORE. Similarly, if all records in a catalog are meant to reflect both the geossDataCore and geossAttribution tags and no other tags, then the catalog is considered homogeneous, and can be registered as GEOSS Data-CORE with attribution. On the other hand, if all records in a catalog are meant to reflect the geossDataCore tag, but only some of the records are meant to reflect the geossUserRegistration tag, then the catalog cannot be considered homogeneous.

One way to achieve an easy and fast metadata updating for each dataset being contributed to the GEOSS Data-CORE is to use homogeneous catalogs. Indeed, only homogeneous catalogs, with respect to the GEOSS Data-CORE metadata tags, can be registered in the CSR to avoid tagging individual metadata records. This is because the CSR, as part of the resource registration process, allows the choice of which of the four metadata tags should be associated with the entire resource being registered. Thus, for example, if a catalog being registered in the CSR will be associated with the geossDataCore tag, then all records in the catalog will be considered as tagged with the geossDataCore tag.

Of course, not all catalogs will be homogeneous with respect to the GEOSS Data-CORE metadata tags. However, one possible way to allow an easy and fast metadata updating if you have a non-homogeneous catalog, is to split the catalog up into two or more homogeneous

catalogs (for the purpose of CSR registration). In this way, each of the homogeneous catalogs can then be registered in the CSR appropriately. This process of splitting requires taking an original catalog and creating one or more additional catalogs so that the metadata records of the original catalog can be placed in the new catalogs in such a way that, in the end, there are catalogs that are homogeneous, as desired.

For example, suppose Catalog A contains 5 records as follows:

Catalog A
1. Copepoda from the coastal upwelling zone of the Chilean Humboldt Current System [cl_udec_copas_copepoda_01]
2. Deep-water and El Niño-related fishes from the northern Humboldt Current System of Chile-Peru [cl_udec_copas_fishes_01]
3. Phytoplankton of the Eastern South Pacific (OBIS, ESPOBIS) [cl_udec_copas_unap_phytoplankton]
4. Deep-water and El Niño-related fishes from the northern Humboldt Current System of Chile-Peru [COPAS_I_Chile]
5. Zooplankton of the Eastern South Pacific (OBIS, ESPOBIS) [zooplancton_copas_cl]

If only records 1 and 3 are GEOSS Data-CORE, then Catalog A is not homogeneous, and prior to CSR registration, the data provider would need to create Catalog B and move records 2, 4, and 5 into it. The final result follows:

Catalog A
1. Copepoda from the coastal upwelling zone of the Chilean Humboldt Current System [cl_udec_copas_copepoda_01]
3. Phytoplankton of the Eastern South Pacific (OBIS, ESPOBIS) [cl_udec_copas_unap_phytoplankton]

Catalog B
2. Deep-water and El Niño-related fishes from the northern Humboldt Current System of Chile-Peru [cl_udec_copas_fishes_01]
4. Deep-water and El Niño-related fishes from the northern Humboldt Current System of Chile-Peru [COPAS_I_Chile]
5. Zooplankton of the Eastern South Pacific (OBIS, ESPOBIS) [zooplancton_copas_cl]

In this case, Catalog A is homogeneous with respect to GEOSS Data-CORE and can be registered as GEOSS Data-CORE. Catalog B can be registered, but will not reflect GEOSS Data-CORE.

Data Discovery and Access

Data registered in the GEOSS can be discovered via the GEO Web Portal:
http://www.geoportal.org/web/guest/geo_home

or via another client of the GEOSS Discovery and Access Broker such as :
http://www.eurogeoss-broker.eu/
(Please note that you will need to run the client using Mozilla Firefox.)

Being able to discover data resources in the GEOSS Data-CORE is an essential first step, and ranking mechanisms are being developed in the GCI to ensure GEOSS Data-CORE resources receive priority in the return of a search. While discovery is necessary, accessing and using the data is the ultimate goal. Data access capabilities are currently under development and will be made available as soon as possible in a separate instruction document.

ANNEX I

Metadata standards, or the organization utilizing the metadata standard, should be consulted as to whether multiple tags should use multiple instances of metadata fields, or whether they can all appear in one metadata field separated by commas, spaces, or some other delimiter. The GCI's ability to recognize the GEOSS Data-CORE tags will succeed in any situation where the tags are placed in a text-searchable metadata field. In the following metadata examples, any expression of a particular tag, such as geossDataCore, can be interchanged by any of the other tags discussed in this document, except where specifically indicated otherwise.

- ISO 19139 metadata
  - This standard is also used by CSW/ISO, OAI-PMH/ISO, GeoNetwork, Deegree, ESRI Geoportal, and SeaDataNet CDI
  - gmd:identificationInfo[1]/*/gmd:resourceConstraints/*/gmd:useLimitation/gco:CharacterString is used
    - \<gco:CharacterString\>geossDataCore\</gco:CharacterString\>
- Dublin Core metadata
  - This standard is also used by CSW/core and OAI-PMH/Dublin Core
  - "Rights" element is used
    - Rights = "geossDataCore"
- OpenSearch
  - RSS response is used with "copyright"element of "channel"
    - \<channel\>\<copyright\>geossDataCore\</copyright\>\</channel\>
  - RSS response is used with Dublin Core namespace and "channel"
    - \<channel\>\<dc:rights\>geossDataCore\</dc:rights\>\</channel\>
  - Atom response is used with "rights" contained in "entry" or "feed"
    - \<feed\>\<rights\>geossDataCore\</rights\>\</feed\>
    - \<entry\>\<rights\>geossDataCore\</rights\>\</entry\>
- WCS, WFS, WMS, WPS metadata
  - Service level "accessConstraints" element in GetCapabilities document is used
    - \<accessConstraints\>geossDataCore\</accessConstraints\>
  - Dataset level uses the ISO 19139 metadata standard as with CSW/ISO
- GBIF metadata
  - Rights element is used
    - \<rights\>geossDataCore\</rights\>
- THREDDS metadata
  - Documentation element, "rights" type is used
    - \<documentation type="rights"\>geossDataCore\</documentation\>
- NetCDF metadata
  - Includes CF and ACDD conventions
  - "license" attribute is used
  - CDL notation as a global attribute
    - license = "geossDataCore"
- DIF metadata
  - Uses "Access_Constraints" for user registration and monetary charge
    - \<Access_Constraints\>geossUserRegistration\</Access_Constraints\>
  - Uses "Use_Constraints" for attribution and GEOSS Data-CORE
    - \<Use_Constraints\>geossDataCore\</Use_Constraints\>