GEO DAB (Discovery and Access Broker): Registration Guidelines

Title	GEO DAB (Discovery and Access Broker): Registration Guidelines
Authors	Stefano Nativi (CNR-IIA), and Mattia Santoro (CNR-IIA)
Date	24 Nov 2015
Document version	4.3
Document type	Draft
Abstract	TBD
Dissemination type	Internal
Document MIME Type	application/pdf
Document language	English
URN	www.uos-firenze.essi-lab.eu/GEO-DAB-Guidelines-document-4.3
URL	TBD
Contact point	stefano.nativi@cnr.it

Outline

Scope and Intended Audience

The main objective of the document is to explain how to contribute datasets to GEOSS by registering directly via the GEO DAB (Discovery and Access Broker).

Data contribution applies the Open Data access policy advocated by the European Commission.

This document contains a set of guidelines that is intended for data and service Providers.

Contact Points

To get further information on the GEOSS Brokering process and this document, please contact:

Stefano Nativi
Mattia Santoro

Department of Florence, Institute on Atmospheric Pollution Research (IIA) of the National Research Council of Italy (CNR)

Introduction

GEOSS and GEOSS Data Core

TBD

GEOSS Providers and the GCI

TBD

Data sharing approach in GEOSS

GEOSS implements a System of Systems (SoS) by applying the broker pattern to realize its central integration and interoperability infrastructure: i.e. the GCI. Figure 1 depicts the simplified SoS components schema.

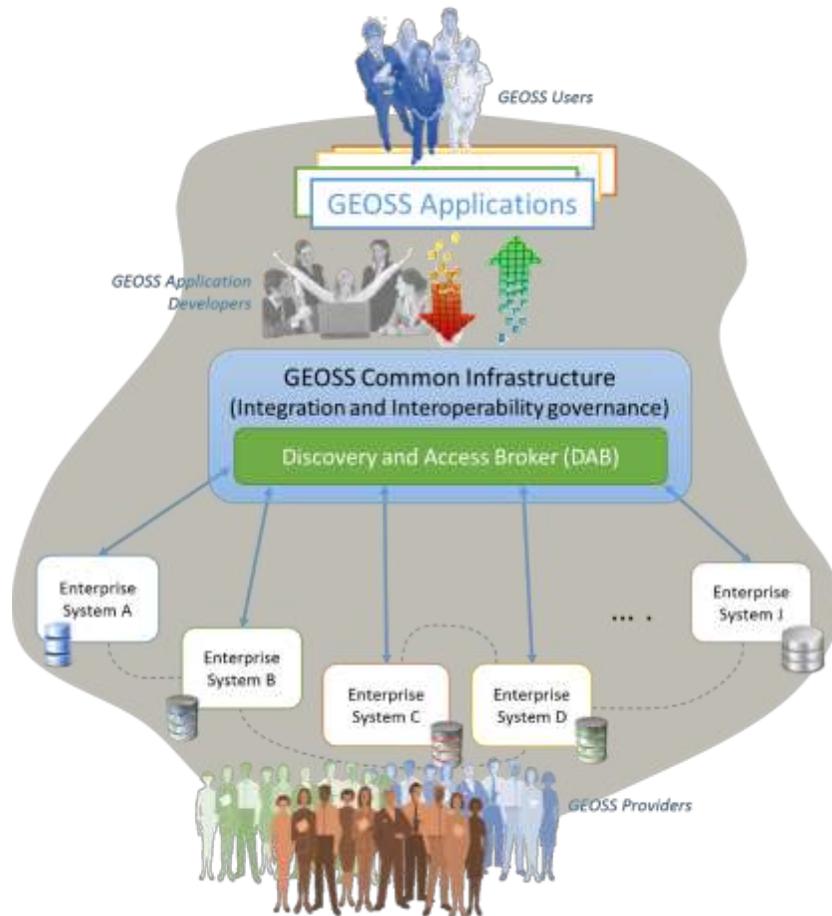*Figure 1. Simplified GEOSS components schema*

GCI adopts a collaborative-acknowledged governance approach. Each GEOSS constituent system, even if it is and remains fully autonomous, is required to follow a set of technical guidelines to enable the DAB to broker it effectively

GEOSS Brokering Process

The GEOSS Brokering process aims to add a new constituent element in GEOSS. This entails:

a) To recognize a new GEOSS Provider;
b) To recognize a new set of datasets –or more than one- shared by the new Provider;
c) To recognize a new enterprise system, which publishes the new set of datasets;
d) To recognize one or more Web/Internet interfaces (or APIs) exposed by the new enterprise system to implement data sharing –i.e. discoverability and access.

The Brokering process starts with a Registration process aiming to get general information on: (a) the Provider Organization; (b) the shared datasets; (c) the enterprise system to be brokered; (d) the interoperability interfaces/APIs published by the enterprise system.

To simplify the registration task, GEOSS decided to apply the expert pattern introducing two different registration phases suited for two different kinds of expertise of the Provider organization (see Figure 2):

- **Administrative registration** –dealing with the Provider Organization and its shared datasets;
- **Interoperability registration (and brokering)** –dealing with the enterprise system and its interoperability interfaces/APIs. Actual Brokering tests are run only in this phase.

The Administrative registration must be the first one to be accomplished. The Interoperability registration includes the tasks necessary for the actual brokering of the Provider's enterprise system.

Actors involved in the Registration and Brokering process

For the scope of the present document, the parts involved in the registration and brokering process are:

- The new GEOSS Provider: hereafter referred as "Provider" –represented with the icon
- The GEO DAB operational team: hereafter referred as "DAB team" –represented with the icon
- The DG RTD of the European Commission: hereafter referred as "EC DG-RTD" –represented with the icon

Registration and Brokering process workflow

The whole process workflow is depicted in Figure 2. Main actor(s) in charge of the single tasks is outlined using different icons. Actors who are not in charge of a task but must be informed about the task outcome are represent using a grayed version of the icon.
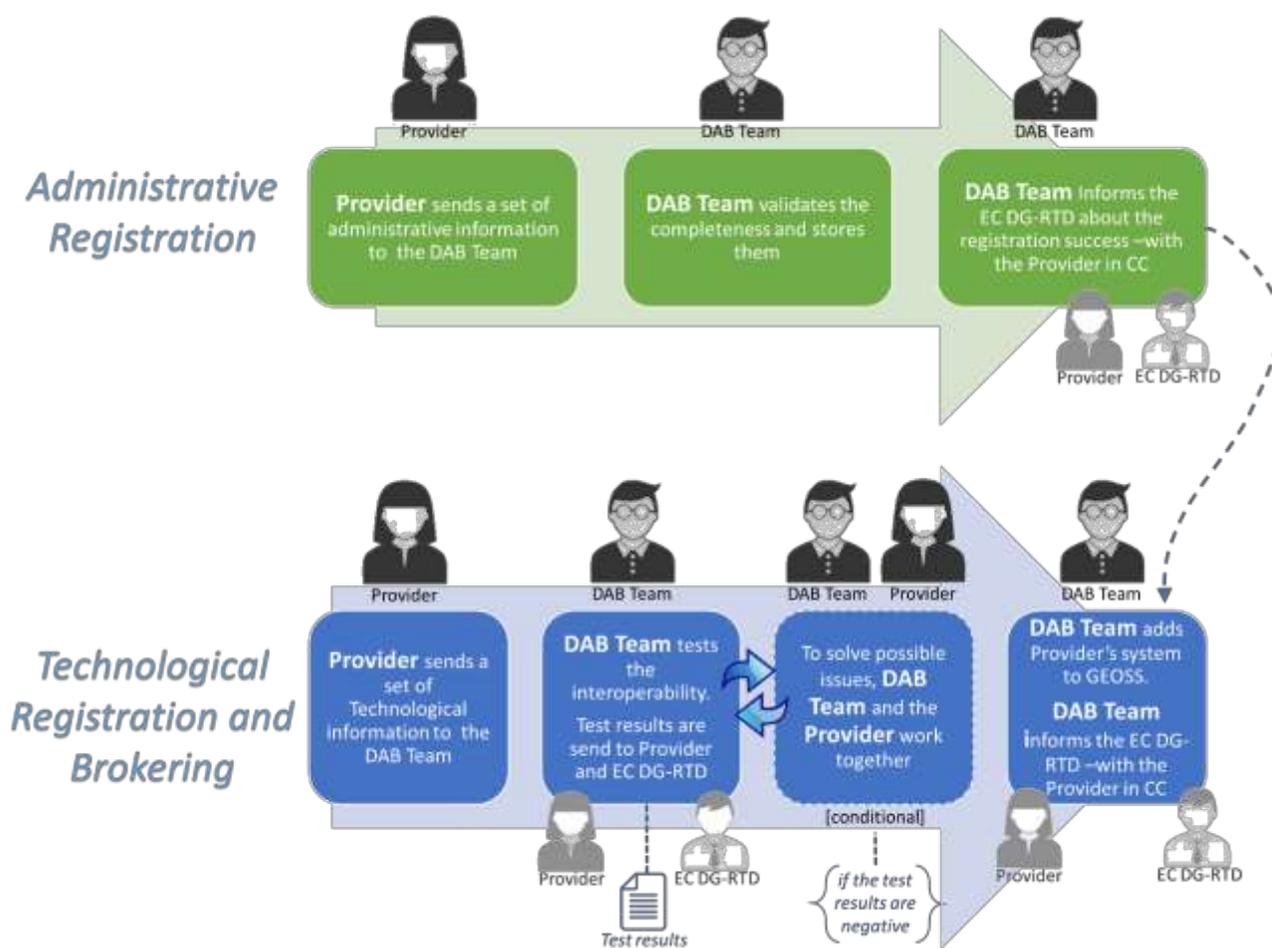
Figure 2. Brokering process workflow

The workflow for a new Provider's system registration and brokering consists of two (almost) parallel sub-workflows:

1. The A**dministrative Registration workflow**, consisting of the following steps:
    a. **Provider** sends a set of administrative information to the DAB Team;
    b. **DAB Team** validates the completeness and format of the received information and stores the in the GEOSS registry;
    c. **DAB Team** Informs the EC DG-RTD –with the Provider in CC
2. The **Interoperability Registration and Brokering workflow**, consisting of the following steps:
    a. **Provider** sends a set of few Technological information to the DAB Team
    b. **DAB Team** tests the Provider's system interoperability with GEOSS. A test results (called interoperability test report) document is generated and sent to the Provider and to EC DG-RTD.
    c. [conditional] If (according to step b.) there exists an interoperability issue, the **DAB Team** and the **Provider** work together to solve it. Then, the workflow returns to step b.

Step b. and c. are cyclically repeated until the interoperability test results are successful.

Note: According to the brokering pattern, the DAB Team will develop any mediation and or transformation module, if required for establishing interoperability.

d. When the interoperability test is successful and the Administrative registration workflow is accomplished, the **DAB Team** can add the Provider's system to GEOSS. **DAB Team** informs the EC DG-RTD –with the Provider in CC.

Both the Administrative registration and the Interoperability registration & brokering workflows must be successfully accomplished to add the Provider's system to GEOSS

Provider's Information required by the registration process

Administrative information

The following <u>seven informative sections</u> must be provided by the Provider to the DAB Team –see step a. of the Administrative Registration workflow. They deal with the Provider organization and the published data:

1. **Organization information**, specifying:
   a. Name
   b. Address
   c. Web site address
   d. Logo
   e. Email address
   f. Telephone number
   g. [optional] Other information
2. **Data policy** applied by the Provider for:
   a. Access policy
   b. Redistribution policy
   c. Use policy
3. **Administrative Contact point**, specifying:
   a. Name and Surname
   b. Role in the Organization
   c. Email address
   d. Telephone number
4. **Technological (Interoperability) Contact point**, specifying:
   a. Name and Surname
   b. Role in the Organization
   c. Email address
   d. Telephone number
5. **Label** to be officially used for the Provider enterprise system in GEOSS.
6. **Data Category**, select one entry from the following list:
   a. "Environmental data"
   b. "Socio-economic data"

7. **Data type(s)**, select one or more entries from the following list:
    a. "Satellite datasets" [condition: *Data Category = Environmental data*]
    b. "In-situ datasets" [condition: *Data Category = Environmental data*]
    c. "Raw data"
    d. "Variable"
    e. "Essential Variable"
        i. for which Community of Practice must be specified
    f. "Indicator"
    g. "Index"
        i. for which Community of Practice must be specified
    h. "Model outcome"
    i. "Other"
        i. an alternative type name must be specified

Technological information

The following five informative sections (note: only one is mandatory; two are conditional and the last two are optional) must be specified by the Provider to the DAB Team –see step a. of the Interoperability Registration workflow. They deal with the Provider's enterprise system and its interfaces:

1. **Web service(s)/Web API(s)** published by the Provider enterprise system to share datasets, specifying:
    (i) service/API **Protocol** implemented by the Web service to understand the data format and the messages syntax . Both REST and SOAP messages are supported –e.g. "OGC CSW 2.0.2", "OpenSearch", "FTP", "OPeNDAP 2.0", "OGC WMS 1.3", etc.
    (ii) service/API **End point** characterizing the specific location for accessing the web service using the specified protocol –e.g. HTTP or FTP URLs.
    (iii) supported **Binding mode** for discovery, select one entry from the following list:
        • Metadata Harvest;
        • Discovery Request distribution.
    (iv) supported **Access mode**, select one entry from the following list:
        • Login required
        • Open
2. Metadata **Re-harvesting time** [conditional: if *Binding mode* = "Metadata Harvest"], the time (in hours) intervening between two harvests operation by the DAB.
3. **Account** to be used to access data [conditional: if *Access mode* = "Login required"], the information of the account specifically created for the DAB.
4. **Number of shared datasets** [optional], the (estimated) number of datasets shared by the system enterprise.
5. **Special interoperability requirements** [optional], any other requirement the Provider wants to specify.

GEO DAB (and the GCI) strongly suggest Providers to support the "Metadata Harvest" mode to improve the discoverability and evaluation of their datasets.

GEO DAB realizes a flexible and extensible platform for enterprise systems mediation and harmonization; it supports the large majority of protocols used by EO Providers to share datasets on the Web –e.g. OGC/ISO protocols, W3C protocols, INSPIRE protocols, TDWG protocols, WMO protocols, OAI protocols, DCMI protocols, UNIDATA/UCAR/NCAR specifications, etc.

Providers are encouraged to implement international standards for interoperability.

Metadata Content

GEO DAB implements discovery and access harmonization for GEOSS.

To achieve that, the datasets, shared by GEOSS Providers, must be characterized by a minimum set of metadata elements.

Required Dataset Metadata

Each dataset contributed by a Provider to GEOSS must be shared along with the following ten metadata elements. Even if only six one are mandatory, it is strongly recommended to provide the optional ones, too:

1. Unique Identification [Optional];
2. **Dataset title** [Mandatory];
3. **Abstract describing the dataset** [Mandatory];
4. **Dataset reference date** [Mandatory];
5. Dataset responsible party [Optional];
6. **Geographic location of the dataset** (by four coordinates or by geographic identifier) [Conditional];
7. **Dataset topic category** [Mandatory];
8. Lineage [Optional];
9. **On-line resource** [Mandatory];

For the semantics of each metadata element, Providers must refer to the ISO 19115 standard specification ("Geographic information — Metadata", 2003 and later corrections).

GEOSS Sustainability clauses

GEOSS sustainability largely depends on the Provider systems sustainability. Therefore, the following clauses are considered mandatory for any GEOSS Provider:

(a) The Provider **shall** promptly inform the DAB Team about any change affecting the information detailed in the Registration and Brokering phases.

(b) DAB Team **shall** promptly inform Providers about any change concerning GEO DAB affecting its interoperability capabilities and/or the GCI architecture.

Annexes

Annex A: Terms and Acronyms definition

For the scope of this document, the following acronyms and terms are defined.

Application Developer	An enterprise developing a GEOSS Application.
Binding	An association between an interface, a concrete protocol, and a data format. A binding specifies the protocol and data format to be used in transmitting messages defined by the associated interface. [*W3C, Web services glossary*]
Broker pattern	The broker architectural pattern is commonly used to structure distributed software systems with decoupled components that interact by remote service invocations. A central broker component is responsible for coordinating communication, such as forwarding requests, returning results and exceptions.
Collaborative-acknowledged governance	Governance strategy applied by GEOSS: constituent systems interact voluntarily with the GCI (via the DAB) to fulfill recognized objectives. However, the constituent systems retain their independent ownership, objectives, funding, and development and sustainment approaches.
Community of practice	A Community of Practice (CoP) is a user-led community of stakeholders, from providers to the final beneficiaries of Earth observation data and information, with a common interest in specific aspects of societal benefits to be realized by GEOSS implementation.
Copernicus system	A possible GEOSS Provider. An enterprise system to be brokered by the GCI through the DAB.
Dataset	Identifiable collection of data [*ISO TC 211*]
EC DG-RTD	The DG RTD of the European Commission, which represents the European Commission in GEO.
End point	An association between a binding and a network address, specified by a URI (e.g. URL), that may be used to communicate with an instance of a service. An end point indicates a specific location for accessing a service using a specific protocol and data format. *[W3C, Web services glossary]*
EO	Earth Observation
Expert pattern	This software engineering pattern recommends to assign a responsibility to who has the information needed to fulfill it.
FTP	FTP (File Transfer Protocol) is a standard network protocol used to transfer files between clients and servers over the Internet.
GCI	GEOSS Common Infrastructure: the GEOSS centralized infrastructure implementing the necessary artifacts for the integration and interoperability of all the GEOSS Providers.
GEO	Group on Earth Observation (https://www.earthobservations.org)
GEO DAB	The GEOSS Discovery and Access Broker (http://www.eurogeoss.eu/broker/Pages/AbouttheEuroGEOSSBroker.aspx)
GEO DAB Team	The CNR-IIA Team configuring and operating the GEO DAB.

GEOSS	The Global Earth Observation System of Systems programme (https://www.earthobservations.org/geoss.php). GEOSS is a SoS.
GEOSS Application	A user-driven application developed on the top of GEOSS data and services.
GEOSS Client	Any human or software system accessing and using the GEOSS content.
GEOSS Data-Core	A distributed pool of documented datasets with full, open and unrestricted access at no more than the cost of reproduction and distribution. The pool is accessible through the GCI.
GEOSS Provider	An organization managing an enterprise System contributing its data and/or services to GEOSS –i.e. one of the constituent system of GEOSS.
GEOSS Users	The human clients of GEOSS.
HTTP	The HyperText Transfer Protocol (HTTP) is the protocol used by the Web to define how messages are formatted and transmitted.
Interoperability test reports	A document reporting the results of the interoperability tests, conducted by the GEO DAB Team, to experiment the interoperability level of a new GEOSS Provider system.
Metadata	Data about data [*ISO TC 211*]
Protocol	A set of formal rules describing how to transmit data, especially across a network. Low level protocols define the electrical and physical standards to be observed, bit- and byte-ordering and the transmission and error detection and correction of the bit stream. High level protocols deal with the data formatting, including the syntax of messages, the terminal to computer dialogue, character sets, sequencing of messages etc. [*W3C, Web services glossary*]
Provider	see GEOSS Provider.
REST(REpresentational State Transfer)	An architectural style, and an approach to communications that is often used in the development of Web services.
Service	an abstract resource that represents a capability of performing tasks that form a coherent functionality from the point of view of provider entities and requester entities. To be used, a service must be realized by a concrete provider system. [*W3C, Web services glossary*]
Service interface	the abstract boundary that a service exposes. It defines the types of messages and the message exchange patterns that are involved in interacting with the service, together with any conditions implied by those messages.
	A logical grouping of operations. An interface represents an abstract service type, independent of transmission protocol and data format. [*W3C, Web services glossary*]
Societal Benefit Area	The Societal Benefit Areas (SBAs) are nine environmental fields (issues) of interest around which the GEOSS is exerting its efforts.
SBA	See "Societal Benefit Area"
SoS	System of Systems (aka supersystem and metasystem) is a large-scale distributed system whose components are heterogeneous enterprise systems that are networked together by a centralized governance mechanism, to accomplish a shared strategic goal.

URL	A uniform resource locator (URL) (aka web address) is a reference to a Web resource, specifying the location of the resource on a computer network and a mechanism for retrieving it –e.g. the HTTP.
Web API (Application Programming Interface)	A development in Web services where emphasis has been moving to simpler REST based communications. RESTful APIs do not require XML-based Web service protocols (SOAP and WSDL) to support their interfaces.
Web Service	A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format. Other systems interact with the Web service in a manner prescribed by its description using REST or SOAP messages, typically conveyed using HTTP.

Annex B: GEOSS useful Information and Recommendations for data Providers

The **GEOSS Data Management principles** to assure a quality service to the Users.

Document URL

The **GEOSS Data Sharing Principles**

Document URL

The **statistical information** about the use of their resources.

Document URL

….

Annex C: Useful publications

Craglia M. [et al.] EuroGEOSS FINAL REPORT [Report]. - [s.l.] : EuroGEOSS Deliverable, 2012.

Edwards A.W., Craglia M. and Nativi S. Advancing the Vision of the Global Earth Observation System of Systems: a European Perspective [Conference] // AGU Fall Meeting Abstracts. - San Francisco (CA) : [s.n.], 2012.

GEO GEOSS Societal Benefit Areas [Online]. - August 14, 2015. - http://earthzine.org/geo-and-geoss-the-group-on-earth-observations-and-the-global-earth-observations-system-of-systems/societal-benefit-areas/.

ISO-TC211 IS 19115 -Geographic information — Metadata [Report]. - Geneva : ISO, 2003.

Khalsa S. [et al.] Brokering for EarthCube Communities: A Roadmap [Online]. - 2013. - August 14, 2015. - http://dx.doi.org/10.7265/N59C6VBC.

Nativi S. [et al.] Big Data challenges in building the Global Earth Observation System of Systems [Journal] // Environmental Modelling & Software. - 2015. - Vol. 68. - pp. 1-26.

Nativi S. [et al.] EarthCube White Paper: The Brokering Approach for Earth Science [Online]. - 2011. - August   14,   2015. -   https://rd-alliance.org/system/files/filedepot/97/earthcube(BrokeringApproach)_v3.3.pdf.

Nativi S., Craglia M. and Pearlman J. Earth Science Infrastructures Interoperability: the Brokering Approach [Journal] // IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing. - 2013. - 3 : Vol. 6. - pp. 1118 -1129.

Nativi S., Craglia M. and Pearlman J. The brokering approach for multidisciplinary interoperability: A position paper [Journal] // International Journal of Spatial Data Infrastructures Research. - 2012. - Vol. 7. - pp. 1-15.

Nativi S., Mazzetti P. and Geller G. N Environmental model access and interoperability: The GEO Model Web initiative [Journal] // Environmental Modelling & Software. - 2013. - Vol. 39. - pp. 214-228.

Nativi Stefano, Mazzetti Paolo and Plag Hans-Peter Towards a Sustainable Geoss. (Global Earth Observation System of Systems) Some Results of the Egida Project [Book]. - [s.l.] : AION, 2014.

Papeschi F., Santoro M. and Nativi S. The GEOSS Discovery and Access Broker APIs [Online] // EuroGEOSS broker. - CNR-IIA, 2014. - August 14, 2015. - http://api.eurogeoss-broker.eu/docs/index.html.