EU Framework Program for Research and Innovation (SC5-18a-2014 - H2020)

Project Nr: 641538

Coordinating an Observation Network of Networks EnCompassing saTellite and IN-situ to fill the Gaps in European Observations

Deliverable D3.3
Mid term revision of the data management plan

Version 1

Due date of deliverable: 1/12/2015
Actual submission date: 29/02/2016

Document control page				
Title	D3.3 Mid term revision of the data management plan			
Creator	CREAF			
Editor	CREAF			
Description	The Data Management Plan (DMP) will be reviewed as requested in http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. The review will be conducted in collaboration with WP1 and WP5. The deliverable will include the outcome of task 3.3.			
Publisher	ConnectinGEOConsortium			
Contributors	ConnectinGEOPartners			
Type	Text			
Format	MS-Word			
Language	EN-GB			
Creation date	01/12/2015			
Version number	0.5			
Version date	26/02/2016			
Last modified by	Ivette Serral			
Rights	Copyright © 2016, ConnectinGEOConsortium			
Dissemination level		CO (confidential, only for members of the consortium)		
	X	PU (public)		
		PP (restricted to other programme participants)		
		RE (restricted to a group specified by the consortium)		
	When restricted, access granted to:			
Nature	X	R (report)		
		P (prototype)		
		D (demonstrator)		
		O (other)		
Review status		Draft	*Where applicable:*	
		WP leader accepted		Accepted by the PTB
		PMB quality controlled		Accepted by the PTB as public document
	X	Coordinator accepted		
Action requested		to be revised by all ConnectinGEO partners		
		for approval of the WP leader		
		for approval of the PMB		
		for approval of the Project Coordinator		
		for approval of the PTB		
Requested deadline				

Revision history			
Version	Date	Modified by	Comments
0.1	01-12-2015	IS_CREAF	Created the basic content of the deliverable
0.2	22-02-2016	HPP_TIWAH	Addition of the SEE-IN content
0.3	23-02-2016	SJ_52N	Addition of the Observation gaps survey results content
0.4	25-02-2016	LM_ARMINES	Review and addition of content
0.5	25-02-2016	IMC_IIASA	Review and addition of content
1	26-02-2016	JM_CREAF	Final review

Contributors	
Acronym	Full name
IS_CREAF	Ivette Serral (CREAF)
JM_CREAF	Joan Masó (CREAF)
HPP_TIWAH	Hans Peter Plag (TIWAH)
IMC_IIASA	Ian Mcallum (IIASA)
LM_ARMINES	Lionel Menard (ARMINES)

Table of Contents

1. Introduction

The purpose of this Data Management Plan (DMP) is to provide an analysis of the main elements of the data management that will be generated by the project.

ConnectinGEO's main mission is not the creation of geospatial datasets but establishing an inventory of data and observations delivered by existing EO networks in Europe. ConnectionGEO also collects relevant standards and the contribution to the GCI standards registry. However, some data can be a result from the Industry challenges.

Having all this information on data, observations and standards, ConnectinGEO will apply a methodology for detecting gaps and redundancies. The result of this methodology is also a data output of the project and will be collected and openly offered as well.

2. Inventory of data sources for ConnectinGEO

2.1. Data set reference and name

ConnectinGEO observation inventory.

2.2. Data set description

In the initial steps of the project, an inventory of observations data sources will be created. The inventory will be created based on GCI Information, DAB, and Copernicus services catalogues that will allow assessing the coherency, continuity, accuracy and accessibility of the information available. It will provide an assessment of the availability and accessibility of data required to derive the information (e.g. EV) or to run the models that generate the forecasts and projections, and it will support a prioritization of the gaps that need to be addressed.

The information and parameters to be gathered and queried for the observation inventory are:

- spatial, temporal, and thematic coverage (considering missing data)
- accuracy/quality (metadata as well as data)
- accessibility (including licenses, standards, and formats)
- cost aspects (e.g. GEOSS Data CORE, fees)
- versions of observation descriptions.

The ConnectinGEO observation inventory data will be complemented with previous inventories analysis and Copernicus services catalogues. A form of access through the DAB will be analysed and implemented if time permits. Efforts will be undertaken to equally consider remote sensing, airborne and in-situ sensors. The new capabilities, introduced by Sentinel missions, will be considered.

The observation inventory will be completed by the description of the observations and measurements automatically (for example, including sensor specifications or hierarchical relations) and with information extracted from scientific literature.

The following list presents the catalogues in the DAB and its number of records:

New Zealand government geodata catalogue	2175
ISPRA Monitoring network	7680
Chile Geoportal	11647
South African Environmental Observation Network	14818
WIS GISC DWD	143879
IGN GeoPortal	50
UK Data Gov (under testing)	6089
FedEO	2551
New Zealand Monitoring Network	32
NIWA data catalogue	188
Geoscience Australia	21301
Red Vial (Road Network, from Ministry of Public Works) -- from Chile	16
HIS Central US	14331907
NOAA Unified Access Framework Catalogue	5114
Limites (Administrative Boundaries, from Ministry of Public Works as well) -- from Chile	5
SeaDataNet	476
EEA SDI Catalogue	414
GMOS Database	819
Global River Discharge Datasets (GRDC/GEOWOW) - Kisters AG	1928
Registered Data	44417
IRIS Event	4163124
IRIS Station	484768
Data Integration and Analysis System (DIAS) - Japan	217
NASA Global Change Master Directory	28108
RASAQM	297
EGASKRO	408
Knossos	536
INPE CDSR	863290
ArcGIS Online ESRI	185000
BYU World Water Data catalog	15
IODE	7509
GBIF	19888012
MEDINA SDI	158
Canadian Monitoring Network	828
CEOS WGISS Integrated Catalogue (CWIC)	1852
Webservice Energy Catalogue	1699
PANGAEA	335877
US Data Gov	85229
One Geology	438

The total number of record is: 40642371.

If an organization is contributing data that is not currently part of the DAB, CNR and GEO together are open to consider adding their metadata to the DAB. In particular, Task 5.4 of ConnectinGEO proposes to do this for the in-situ networks that are found not to be part of GEOSS yet. ConnectinGEO will collaborate with them on their integration.

2.3. Standards and metadata

Metadata on accuracy/quality will be collected about the data included in the observation inventory. It will originate from current information in the metadata concentrated in the DAB and the GCI.

Metadata explaining the network source of the observations will be also collected. We will limit ourselves to the following fields mainly extracted from the ISO 19115 core metadata:

- Id
- Title [*Name of the network*]
- Abstract
- Producer
- Topic category (1..*)
- Theme/topic (0..*) [*SBA main domain*]
- Measurement (0..*)
- Keywords (0..*)
- EV (0..*)
- Spatial extent (0..1) [*Spatial scale of the action of the network*]
- Temporal extent (0..1) [*Temporal scale of the action of the network*]
- Resolution/scale (0..1)
- Series/parent resource (0..1)
- Link to data (0..*)

Extra requirements can be the following:

- Location of the action of the network
- Administrative aggregation level
- Date of creation
- Availability of background information (report, raw data, metadata)
- Concrete tools / data / guidelines published
- Comments

2.4. Data sharing

As mentioned in the previous section, the data collected in the observation inventory will be integrated during the project. This database will feed the ConnectinGEO methodology on gap and redundancies analysis.

The access to this information will be public via the project website. No registration will be required to access the database of the inventory. The information will be available for consulting and for downloading.

The observation inventory requirements, the database schema and the queryable fields, will be also publicly available to the inventory users as a way to better understand its database and the design.

The database will be then processed by the ConnectinGEO team and statistical analysis of the observation inventory database will be elaborated and made public with the actual data.

2.5. Archiving and preservation (including storage and backup)

The long-term preservation of this dataset is strongly related to the long-term preservation of the inventory. The inventory will be preserved by CNR, the DAB responsible party: CNR. If it's found useful, the inventory can be updated in the future following the same project methodology.

There is no preservation strategy for the DAB content due to the regular and constant update of the DAB. This way of preserving the inventory as a snapshot of the DAB current status is important. After the end of the project, it is reasonable to assume that the preservation of the inventory will be related with the persistence of the DAB as a key component for the GEOSS infrastructure. The cost and efforts of preserving the database will be assumed by the CNR.

3. Inventory of user needs from the Socio-Economic and Environmental Information Needs Knowledge Base (SEE IN KB)

3.1. Data set reference and name

Inventory of user needs from Socio-Economic and Environmental Information Needs Knowledge Base (SEE IN KB).

3.2. Data set description

The GEOSS User Requirements Registry (URR) was the primary place to collect information on Earth observations requirements and the application and users that depended on these observations. A Ministerial guidance provided to GEO during the Ministerial Summit in 2014 requested that GEO "develop a comprehensive interdisciplinary knowledge base defining and documenting observations needed for all disciplines and facilitate availability and accessibility of these observations to user communities." In response to this guidance, the URR has been transformed and extended into the SEE-IN KB.

The contents of the SEE IN KB reflect the knowledge, information, and data needs of a wide range of applications and users. The data model of the SEE-IN KB is based on unstructured objects (Figure 1). Each entity in the SEE-IN KB is an object. Objects can belong to one or more groups (for examples, see next paragraph). A group is defined by a set of attributes selected from a master set of attributes (the ontology). Examples of attributes are Variable, Resolution, Latency, Accuracy, etc. Each attribute comes with a specified domain taken from a master set of domains. Examples of domains are an interval of real number, a list of variables, or a list of specific terms. Groups also are associated with rules for how members of the group can act and interact with other objects. A special group are Links (which are also objects). Links define the relationship between two objects in the same or two different groups. This link concept is used to capture connectivity. With the group concept, new groups can be introduced as needed.

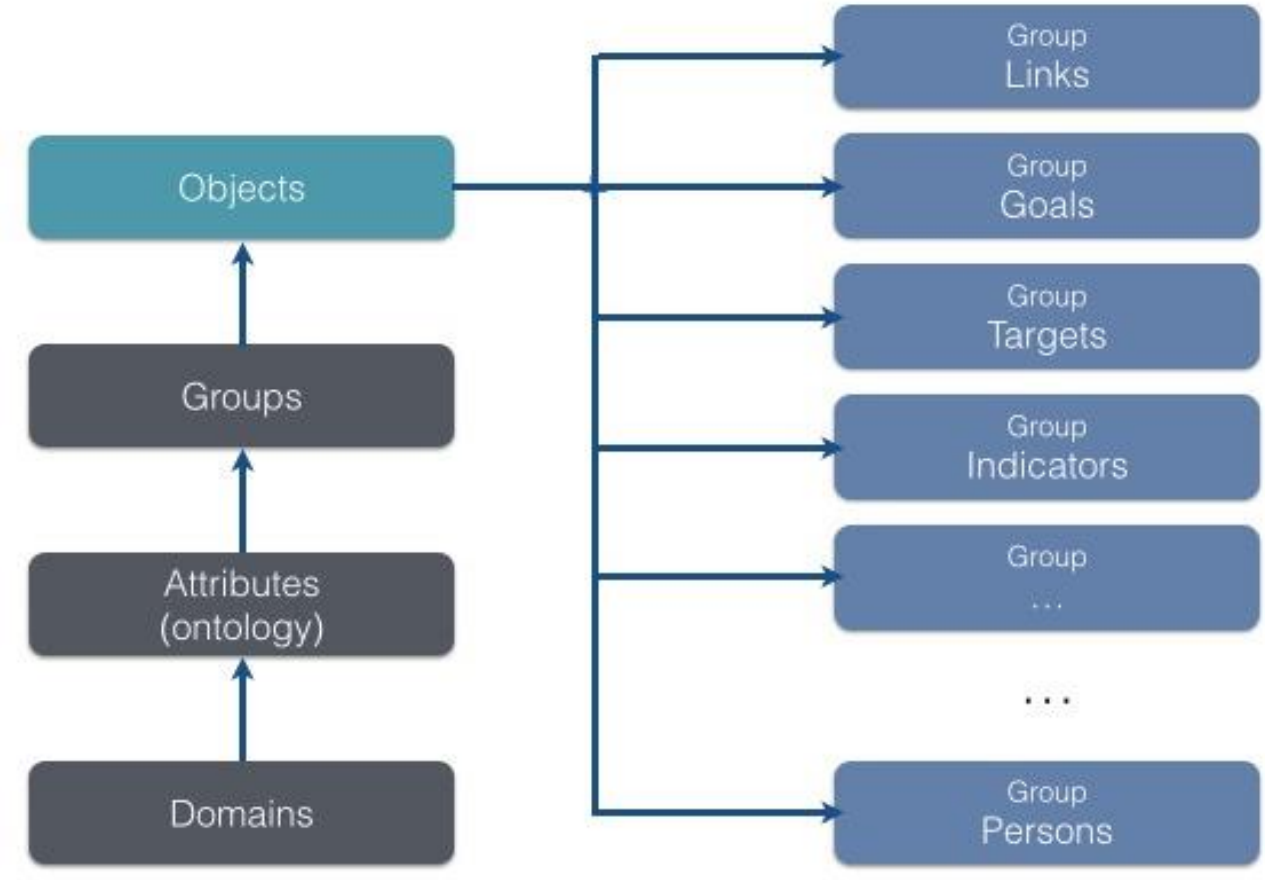


Figure 1. Main elements of the data model of the SEE-IN KB. The dark-gray boxes define the domains for attributes, the ontology, and the known groups. Each entity in the SEE-IN KB is an object, which are registered in an inventory, where the entity is associated with one or more groups. Details on the object are provided in the body of object, depending on the group membership. The objects in the Link group capture the connectivity and relationships between object.

The groups inherited from the URR include User Types, Applications, Requirements, Research Needs, Infrastructure Needs, Technology Needs, Capacity Needs, and Links. Additional groups include, among others, Persons, Models, Services, Datasets, Essential Variables, Variables, and several groups for various types of gaps.

The concept of Essential Variables (EVs) is implemented. Based on the working in ConnectinGEO WP2, rules are made available to link societal goals (such as the Sustainable Development Goals, SDGs) to EVs.

Existing set of EVs can be linked to societal goals and benefits. The results of the review of EV developments in the GEO SBAs will be published in the SEE-IN KB.

The SEE-IN KB serves as the primary source for “user needs” in ConnectinGEO and will be utilized for at least one of the gap analysis approaches in the ConnectinGEO methodology. Moreover, the outcomes of the ConnectinGEO gap analysis will be published in the SEE-IN KB. The information in the SEE-IN KB will not be part of the Observation inventory but will be used later on in the project to compare offerings (data) with needs and determine the gaps. Also it will be used to set priorities.

3.3. Standards and metadata

The catalogue and metadata standards are applied to the SEE-IN KB to the extent applicable.

3.4. Data sharing

The SEE-IN KB is open access to the world-wide community. Guests can view and search all contents. Users can sign up and request editing authority. The SEE-IN KB has several options for user to publish objects in all available groups, establish new groups and develop rules for queries and analyses.

3.5. Archiving and preservation (including storage and backup)

The SEE-IN KB is developed as open-source software with community participation. It is a constantly evolving knowledge base and both information contents and rules can be updated through web-based interfaces. No efforts for preserving the current status of this knowledge base will be made in ConnectinGEO. The gap analysis will be conducted by ConnectinGEO in real time.

4. Observation gaps survey results

4.1. Data set reference and name

Observation gaps survey results.

4.2. Data set description

ConnectinGEO aims to facilitate a broader and more accessible knowledge base to support the needs of the GEO SBAs and their users.
To better detect gaps in the accessibility of observation data, an online survey to collect feedback from scientists and potential users has been developed. The idea of this survey is to help identify existing issues with data availability, technology, data policies and data quality which need to be addressed to improve and support the work in different SBAs.

This survey has been created and conducted using Google Forms. The results of this survey will be exported for further processing while the original survey will be removed from the server.

The survey is available via this URL: http://goo.gl/forms/qI7JvS6id0

4.3. Standards and metadata

The data will be stored by Google Forms in a simple tabular form. Thus, no specific standards will be used. However, the export of the data will be performed using a CSV (comma separated values) structure.

4.4. Data sharing

The results of the survey will be published in an aggregated form as part of a ConnectinGEO deliverable (PDF format). As the survey was structured with a low entry

barrier to ensure a higher number of responses, the participants are not requested to allow the publication of their answers. Thus, for data privacy reasons, the raw data of the survey cannot be published. However, depending on the interest of external users, certain results of the query may be made accessible in anonymous form.

4.5. Archiving and preservation (including storage and backup)

The report will be stored on the web site, by the project coordinator (CREAF) and by 52°North. The raw response data of the query will be stored by 52°North.

5. Observation inventory feedback

5.1. Data set reference and name

Observation inventory feedback.

5.2. Data set description

A publicly accessible feedback catalogue integrated into the observation inventory coming from an extension of the GeoViQua feedback model will be set for geospatial data providing feedback for missing data, i.e. without a reference to a dataset.

5.3. Standards and metadata

Extension of the GeoViQua feedback model for geospatial data with the possibility to provide feedback for missing data, i.e. without a reference to a dataset will be used.

5.4. Data sharing

A single sign-on system in the feedback catalogue (and crowd-sourcing) while trying to be as easy to use as possible, for example using authentication protocols such as OAuth, SIML 2.0. Connect with the planned GEOSS single sign-on system.

User feedback data will be made available featuring no information of the identity of the people providing it.

5.5. Archiving and preservation (including storage and backup)

The Observation inventory feedback is a constantly evolving database that can be updated form the web. No efforts for preserving the current status of this dataset will be made in ConnectinGEO as the Observation inventory feedback will continuously evolve during the project. At the end of the project a copy of its content will be saved and preserved by CREAF.

6. Technical and support documentation about the use of the ConnectinGEO methodology

6.1. Data set reference and name

Technical and support documentation about the use of the ConnectinGEO methodology book.

6.2. Data set description

The ConnectinGEO methodology to identify and assess the priority of gaps will be published as a book.

The book will cover the different approaches that ConnectinGEO will put in practice:

1. Top-Down thread 1: Identification of a **collection of observation requirements and specifications** from **generic goals** for sustainability of the global civilization as expressed in the GEOSS Strategic Targets, the SDGs, and the adherence to the planetary boundaries.
2. Top-Down thread 2: Incorporation of material from international programs such as Future Earth, Belmont Forum, the Research Data Alliance and community assessments of socio-economic benefits of Earth observations.
3. Bottom-up thread 1: A **consultation process** in the current EO networks, consisting of collaboration platforms, surveys and discussions at workshops and even involvement of citizen science.
4. Bottom-up thread 2: A careful **analysis of the observations and measurements** that are **currently** available in GEOSS Discovery and Access Broker complemented by other means (e.g. scientific literature).
5. Bottom-up thread 3: The realization of a series of real **industry-driven challenges** to assess the problems and gaps emerging during the creation of business opportunities.

The book will also cover the results of applying the methodology into GEOSS

6.3. Standards and metadata

The book will be available online in on PDF format online and in paper.

6.4. Data sharing

The book will be freely available on the internet and will be distributed on GEO events.

6.5. Archiving and preservation (including storage and backup)

The book will be available in the ConnectinGEO website and CREAF will preserve it.

7. Information regarding the Essential Variables

7.1. Data set reference and name

Information regarding the Essential Variables.

7.2. Data set description

The thesauri of EV and the indicators to retrieve them. One of the conclusions of the Bari Workshop was that different communities have different levels of maturity for their EV. We can classify EV communities can be classified into:

- Communities that have defined their EV
- Communities that have EV candidates but consensus process is under way
- Communities that do not have EV but there are some obvious candidates
- Communities that have not considered the EV or EV cannot be applied

7.3. Standards and metadata

This data will be provided as a PDF report.

7.4. Data sharing

A list of final agreed EV will be provided as well as the indicators to retrieve them.

7.5. Archiving and preservation (including storage and backup)

The report will be made accessible on the ConnectinGEO website and CREAF will assume the responsibility to preserve the report.

8. ENEON composition and EO network panorama

8.1. Data set reference and name

ENEON composition and EO network panorama.

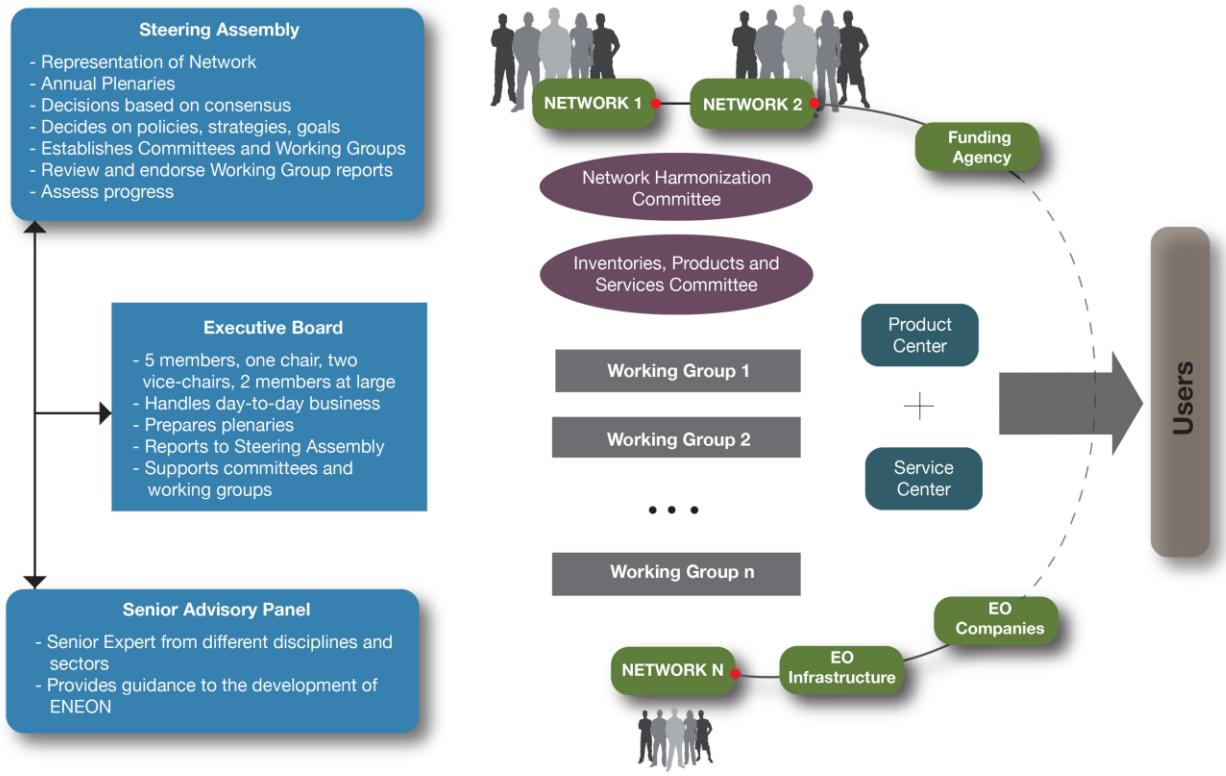


Figure 2. ENEON Structure

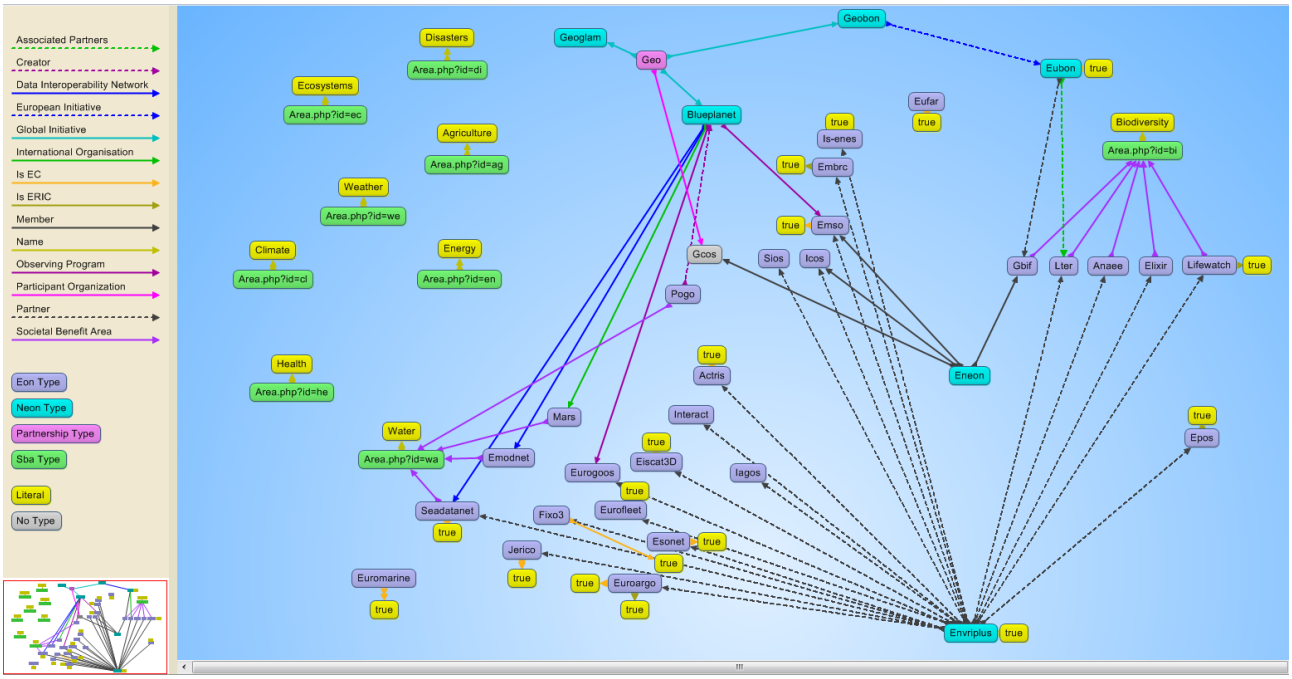
8.2. Data set description

The activities of the ENEON in the project will have as a main goal: to feed a consultation process that will be complemented by a systematic analysis of the available data and metadata to draw a coherent picture of the variety of used data APIs, policies and indicators.

A list of EO networks in Europe collecting:

- A list of models that are being used for forecasting and projecting environmental conditions.
- A list of their data and observation needs.

Due to the intricate relations between networks, the data model will adopt the RDF model.



RDF schema of some of the EO Networks in Europe

VGI systems data will be also considered as can supplement traditional in-situ data collection.

8.3. Standards and metadata

RDF model and a triple store will be used as a way to store the data. We will share this information in the ENEON as an RDF turtle format.

8.4. Data sharing

We will share the data of the ENEON composition and EO network panorama freely and openly on internet.

http://www.eneon.net/

8.5. Archiving and preservation (including storage and backup)

The ENEON composition and EO network panorama triple store will be kept by CREAF that will assume the responsibility to preserve it.

9. Solar renewable energy industrial challenge data

9.1. Data set reference and name

Solar renewable energy industrial challenge data.

9.2. Data set description

Decision makers and project developers of solar power plants (photovoltaic, solar concentrating technologies) have identified the Surface Solar Irradiance (SSI) and its components as an EV for their business development. SSI observations are crucial in the process of selecting the most suitable location of new plants but there is only a sparse network of in-situ pyranometric stations. In the first phase satellite or numeric weather model data are used as a global location identification solution. In a second phase in-situ sensors are deployed for periods of at least one year to collect data, which are more detailed and can be the base for solid business models, which are crucial to acquire funding.

Nevertheless networks of in-situ pyranometric sensors, satellites, or numeric weather models are not able to fully answer the need of industrials and project developers if they are taken separately, but solutions can be found in the conjunction of these EO systems. Still missing is the means to exchange in-situ measurements across companies and between stakeholders in the market which could considerably improve business opportunities. In this task we will provide an open solution for interoperable exchange of SSI data comprising in-situ time-series observations as well as sensor descriptions. The practical solution will be based on the experience from other domains.

It will allow a comprehensive understanding of gaps in capacity and generally, more important, bridge the industrial and the research communities in GEO.

Several companies will use a dedicated platform with pre-existing in-situ measurements for test but also with the important possibility to share their own legacy in-situ measurements to be part of this use case.

9.3. Standards and metadata

The platform will be integrated as a new component into an existing Spatial Data Infrastructure (SDI) dedicated to Energy and Environment namely the already existing and active Webservice-Energy.org GEOSS Community Portal. A key component of this Community Portal is the OGC Catalogue providing access to more than one thousand energy and environment datasets through interoperable Web-services. This catalogue is already registered in the GEOSS Common Infrastructure (GCI) and is harvested on a weekly basis by the GEOSS Discovery and Access Broker (DAB). Consequently the content of the catalogue is available for GEOSS user on the Geo Web Portal (GWP). The in-situ measurements that will be deployed as form of sensor Web-services will be registered as metadata in the Webservice-Energy catalogue. Thus, the full scope from search and discovery of existing resources covered by the catalogue to the selection, display and download of measurement of interest from the dedicated sensor platform will be taken into account. The ability to provide the users with measurement download is a key aspect to start engaging the Energy community to share, release and use in-situ measurements. The participants will provide their feedback on the benefit of such a platform and on the networking of resources for improving their daily work and increasing their business. Therefore this challenge can be a great tool to measure the impact of the ConnectinGEO methodology approach in a specific community and use case.

Regarding gaps in data models, standardized data and metadata models will be designed suitable for in-situ SSI observations as profiles of OGC Sensor Web Enablement standards. Profiles for XML encodings are defined as Schematron rules. The data and metadata models must take into account spatiotemporal coverage, lineage, data quality, IPR, the existing terminology and units of measure.

Regarding gaps in in-situ measurements, the catalogue platform with data management functions will be extend based on the new (meta)data profiles and the GEOSS recommendations for interoperability based on 52°North Sensor Web components. The platform will allow to

- (i) visualise sensor locations on a map,
- (ii) visualise measurements as time series plots and in tabular form,
- (iii) display sensor metadata at different levels of detail,
- (iv) upload of data by data owners, and
- (v) download raw observation data for offline processing. It will support company to prepare, pre-process and integrate their datasets.

9.4. Data sharing

An extension of the existing portal for the energy community allows access to data (particularly in-situ) instead of merely metadata and to allow data sharing between companies and organisations. This will put to the test the capacity of cooperation in the SSI community by introducing an unprecedented level of collaboration and eventually help to detect gaps in European EO networks. The pilot deployment will help to set priorities in the energy domain. It will also explore the possibilities to work together with weather/climate networks.

9.5. Archiving and preservation (including storage and backup)

ARMINES is the institution to store and preserve the data after the project end.