

# ARCHITECTURE OF SPATIAL DATA INFRASTRUCTURE (SDI) (DRAFT)

Author: Carlo Cipolloni  
September 20, 2018  
Version 1.0

## **A SDI platform to support the SEIS in the ENP South region**

### ***Introduction***

Based on the Contribution Agreement signed between UN Environment Programme/MAP and European Environmental Agency (B03437/B2016/R0/ENIS/EEA.56593) which entered into force on 22 September 2016, UN Environment Programme/MAP and EEA agreed to cooperate for the implementation of the EU Project “SEIS-South Support Mechanism”. And in view of improving the availability and access to environmental information to the benefit of effective and knowledge-based policy making in the ENP South region and ensuring proper linkages with their respective work programmes.

One of the objective of this agreement is to support the implementation of the ‘SEIS Support Mechanism-South’ in the field of ‘Infrastructure and data management’, more specifically to: (i) support the countries in enhancing their national infrastructure and data systems and (ii) adjust and extend as appropriate the regional data infrastructure and management systems to cover the data management for assessing the state of environment of the Mediterranean sea based on the agreed H2020/NAP Indicators<sup>1</sup>.

Development of the current SDI platform further builds up on the Regional Analysis of Data Flow Management in Terms of Infrastructure<sup>2</sup> prepare by UN Environment/MAP Info RAC and UN Environment/MAP MED POL. This analysis was carried out on the basis of the answers provided by the countries to the online questionnaire concerning the management of the data flow in terms of infrastructures in the framework of the ENI SEIS II project. The analysis focused on information on public access to data, data format, data policy and the presence of an SDI in countries.

Despite UN Environment/MAP Barcelona Convention has well developed reporting tools for the countries to fulfill their reporting obligations, the platform would facilitate the reporting of the countries, by allowing them to share in collaborative manner the basic dataset and metadata needed to support the Barcelona Convention reporting.

Info-RAC has developed an SDI platform named **InfoMapNode**. The main target of the SDI (Spatial Data Infrastructure) platform is to create a suitable support to the south Mediterranean countries for providing dataset and metadata in a standard way, in order to create a better quality of data and improve information management for policy makers in the Mediterranean sea.

### ***SDI Definition and scope***

A Spatial Data Infrastructure (SDI) is defined mainly as a framework where specific policies on metadata and data sharing and use, are based on institutional arrangements, technologies, and people that enables the sharing and effective usage of geographic information by standardizing formats and protocols for access and interoperability.

<sup>1</sup> Report of the Meeting, 2nd ENI SEIS II South Support Mechanism Regional Workshop on Indicators Athens (Greece), 17-18 April 2018.

<sup>2</sup> Info. Document: Regional Analysis of Data Flow Management in Terms of Infrastructure, 2nd ENI SEIS II South Support Mechanism Regional Workshop on Indicators Athens (Greece), 17-18 April 2018.

The growing need to organize data across the Mediterranean area and the need to create a system or environment platform to support participants pushes to have an SDI.

SDI is the Infrastructure to create an environment that will enable a wide variety of users to access, retrieve, and disseminate spatial data in an easy and secure way. In principle, SDIs allow the widespread sharing of data, which are extremely useful as it allows the users to save resources, time, and effort in acquiring new sets of data, avoiding duplication of expenses associated with generation and maintenance of data as well as their integration with other datasets.

For that reason the main targets of SDI are to:

- Increase available of spatial data and their usability;
- Reduce costs of geographic information while making geographic data more accessible and wider format (possible in open licence);
- Reduce duplication of efforts among the Public Authorities;
- Increase partnerships between states, counties, cities, academia, and the private sector.

In other words, SDI should be seen as part of wider e-Government initiatives, like the European Union (EU) directive on Infrastructure for Spatial Information in the European (INSPIRE Directive).

The ENI SEIS South II project further builds on the Decision IG.21/3<sup>3</sup>, COP 18, Istanbul, regarding data management in Mediterranean by developing a Shared Environmental Information System (SEIS) supporting the regular production and sharing of quality assessed environmental data, indicators and information. Therefore the SDI represents the core element to support the share system.

Based on SDI the participants country to the ENIS SESI South II project can establish a spatial data infrastructure via the Internet that facilitates the sharing of geographic information in a standardized way, following specific technical and nontechnical issues, ranging from standards, organizational and procedural issues, and data policies, to the creation and maintenance of web services.

The SDI is composed by a set of technical and non-technical components that facilitate the sharing of geographic information, we can identify four main domains related to SDI (fig. 1): Data, Metadata, Network services and share agreements.

---

<sup>3</sup> 18th Ordinary Meeting of the Contracting Parties to the Convention for the Protection of the Marine Environment and the Coastal Region of the Mediterranean and its Protocols Istanbul (Turkey), 3-6 December 2013. Annex II: Ecosystems Approach including adopting definitions of Good Environmental Status (GES) and targets.

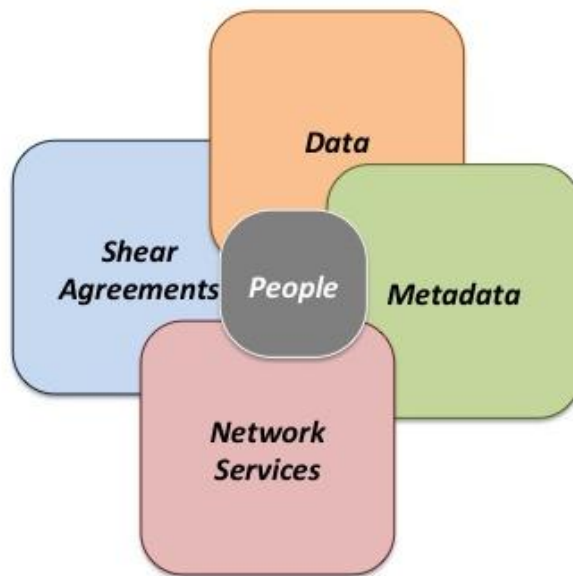


Figure 1 The SDI General domain.

With the exception of the share agreements that is only legal requirements and it not provides technical guideline for the Infrastructure, the other domains provide technical and sometime legal requirements how to build in a correct way the Infrastructure.

More specifically, the Metadata component is composed by software application that provide a system to compile or charge metadata in a standard template and also define the rules how to fill in the template. In this domain should be considered the database and software components, where metadata are stored and collected and the application to provide network service to access to this information or to federate similar catalogue from other Organizations. The Data component involves different elements related to the data store system, normally the database software and applications to manage and ingest these, as well as the standardization, harmonization and validation processes tools that permit to have an high semantic interoperability, a quality check and an integration on data.

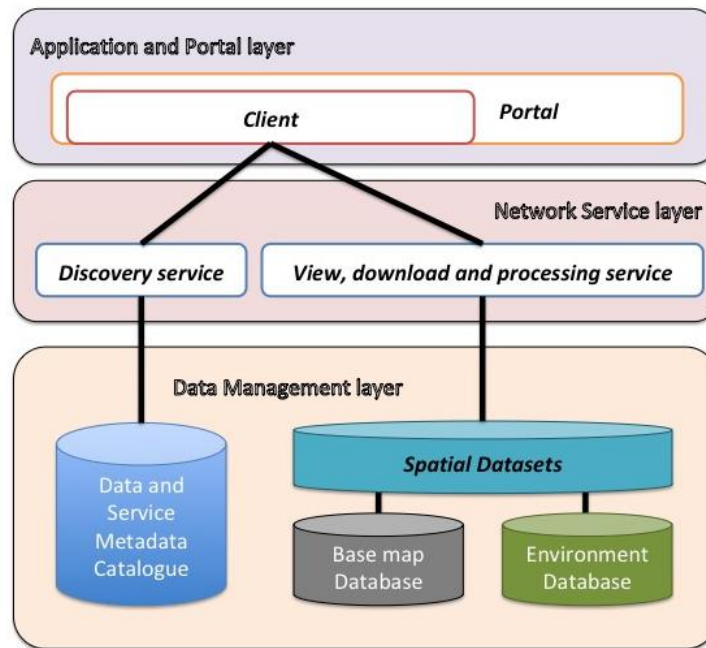
The Network service component represents the software applications and rules to make data, metadata and information in general approach available, accessible and interoperable in different systems. The technical aspect is addressed as software configuration, as standard encoding and format models used.

These sets of components that represent the SDI, ensure some main functionality for the user, as:

- Data Management and geocoding;
- Searching and view of data and metadata;
- Downloading datasets (in accord with their data policy);
- Publish and share documents, maps and layers.

A schematic view of architecture about the technical and software standard composition of SDI is provided in figure 2.

In the next section will describe more in detail the INFO/RAC spatial infrastructure to support ENI SEIS South II project and the orchestration of each components and their role.



*Figure 2. General schema of a Spatial Data Infrastructure*

### ***INFO/RAC Spatial Data Infrastructure for ENI SEIS South II***

Based on the general architecture described in previous section the Information and Communication Regional Action Centre (INFO/RAC) has set-up and customised their infrastructure to host a specific SDI to support the project. This Infrastructure is built to support all the participants in sharing and collecting data, but nevertheless the SDI components could be replicated by the single participant countries in order to create a federate system that facilitates the data sharing.

According with this vision we can propose three different models to collect and share data in the ENI SEIS domain, these are also different models to support each participants in data management process (the three model are also explained graphically in figure 3):

- Model A: A country haven't SDI and do not want to replicate it in the Organization or to require for security reasons to not federate own system with the central INFO/RAC node, can use only the central one developed by INFO/RAC.
- Model B: A country has a SDI, but it's not at all developed to be interoperable with central node and it wants also to improve its technological components, or it wants to build a new one. In this case it can replicate one or more software part and after a configuration procedure, it can federate its SDI with the central one.
- Model C: A country has a SDI well organized and also interoperable in according with the international standard; it can federate directly with central node of INFO/RAC.

Implementation of the Shared Environmental Information System (SEIS) principles and practices in the ENP South region – SEIS Support Mechanism (ENI SEIS II South)

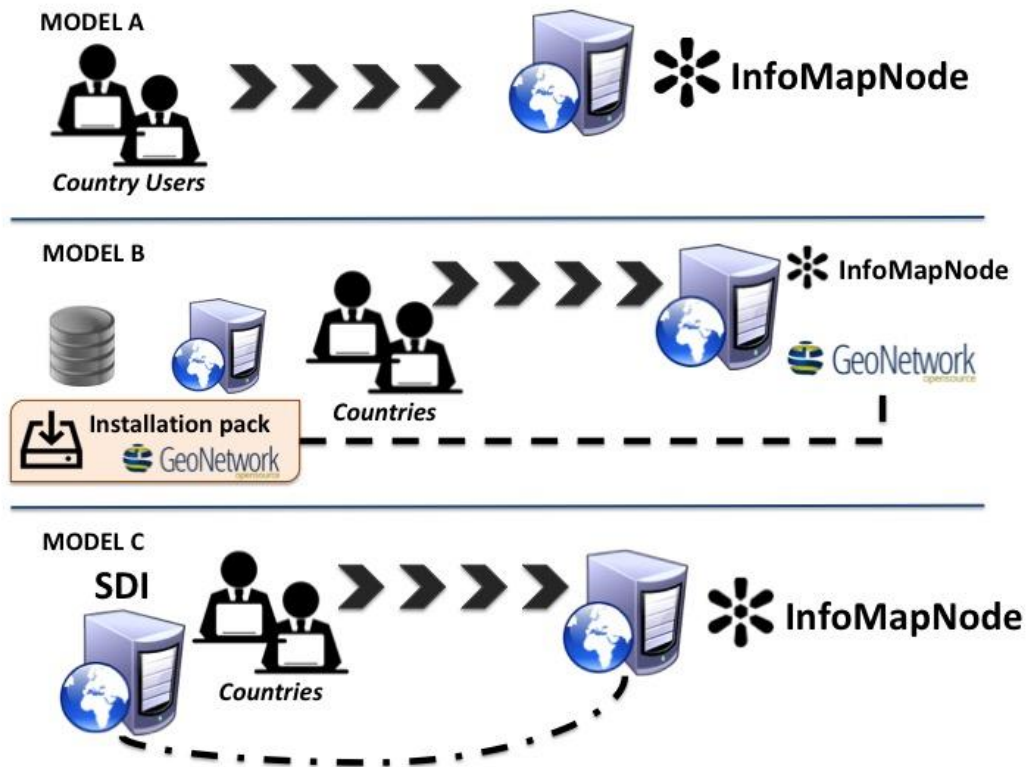


Figure 3. Interconnection models to collect and share data and information among the SDI.

The INFO/RAC has built up its Spatial Data Infrastructure (*InfoMapNode* – figure 4) mainly using open source software and at the same time following the international standards in order to be compatible and interoperable with all the other platforms (figure 5).

The main scope of *InfoMapNode* platform is to support users to collect and share data in a simple manner and the main goals are: share data without being a GIS expert, simplify search, view and download the data, manage the ancillary documents, manage and share in a simple way maps and layer also with a simplified style system, manage the users in different groups in order to support a fine grain access control to support different data policy of south Mediterranean countries.

The main components, following the general architecture of SDI defined before, are:

- Spatial Data Management system to store, query and manage directly the data in the database.
- A Catalogue system with a specific discovery service to harvest, search and query metadata and with integrated Metadata editor;
- Network Service system based where data are distributed and elaborated via Web Service (mainly following OGC Standard);
- A portal within a client to search, view, query and analyse the spatial data;



## Implementation of the Shared Environmental Information System (SEIS) principles and practices in the ENP South region – SEIS Support Mechanism (ENI SEIS II South)

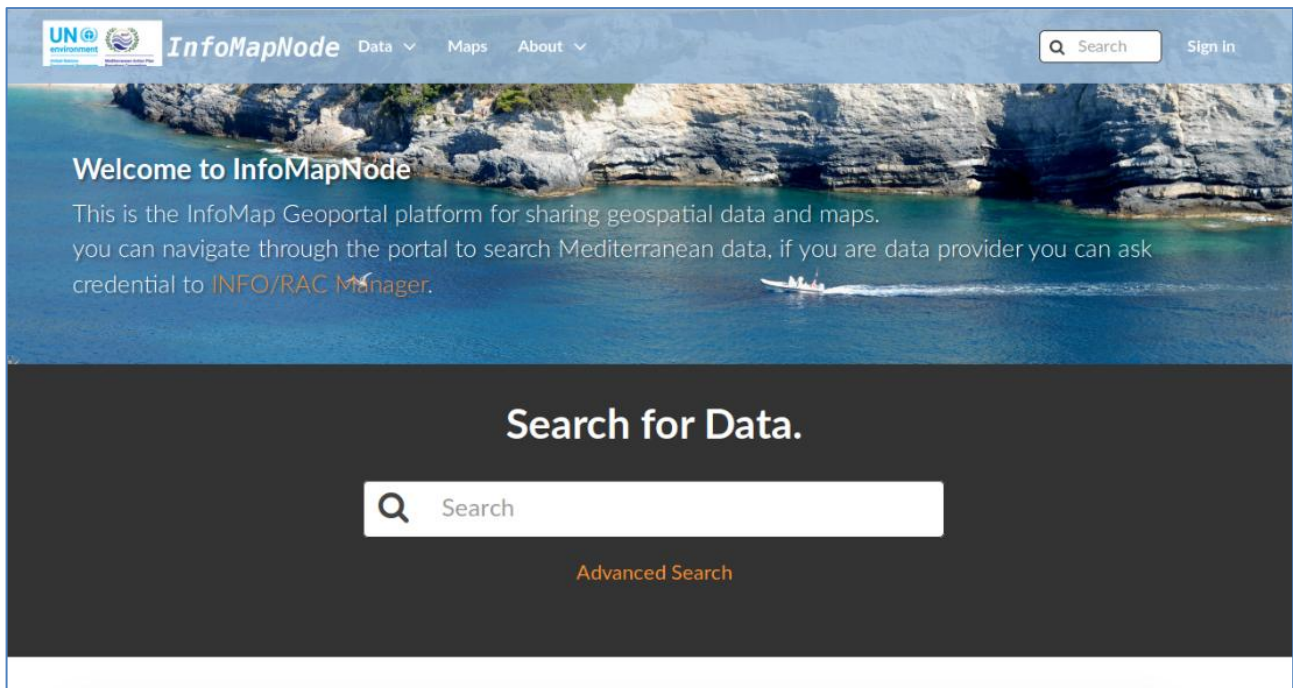


Figure 4. The InfoMapNode platform

### Spatial Data Management system:

This component is the Database software based on *PotgreSQL* and *POSTGIS* as spatial extension to support geographical features. Both software are open source software and as Database Management system can used with GIS software, like *QGis* or a Database Administration Tool like *pgAdmin* (both are open source software as well).

The InfoMapNode platform provides also an opportunity to charge directly data and geographical feature in the PotgreSQL store.

### Catalogue system:

The catalogue itself is an instance within the database but with specific data type to manage the content of metadata files. The software to manage and edit the metadata is embedded in the GeoNode component (it will be described better letter on this section), but in order to support in the SDI the possibility to harvest different metadata catalogue services also in OGC CSW 2.02 format from different countries and metadata hub, the system is implemented on the open software component *GeoNetwork*.

### Network Service system:

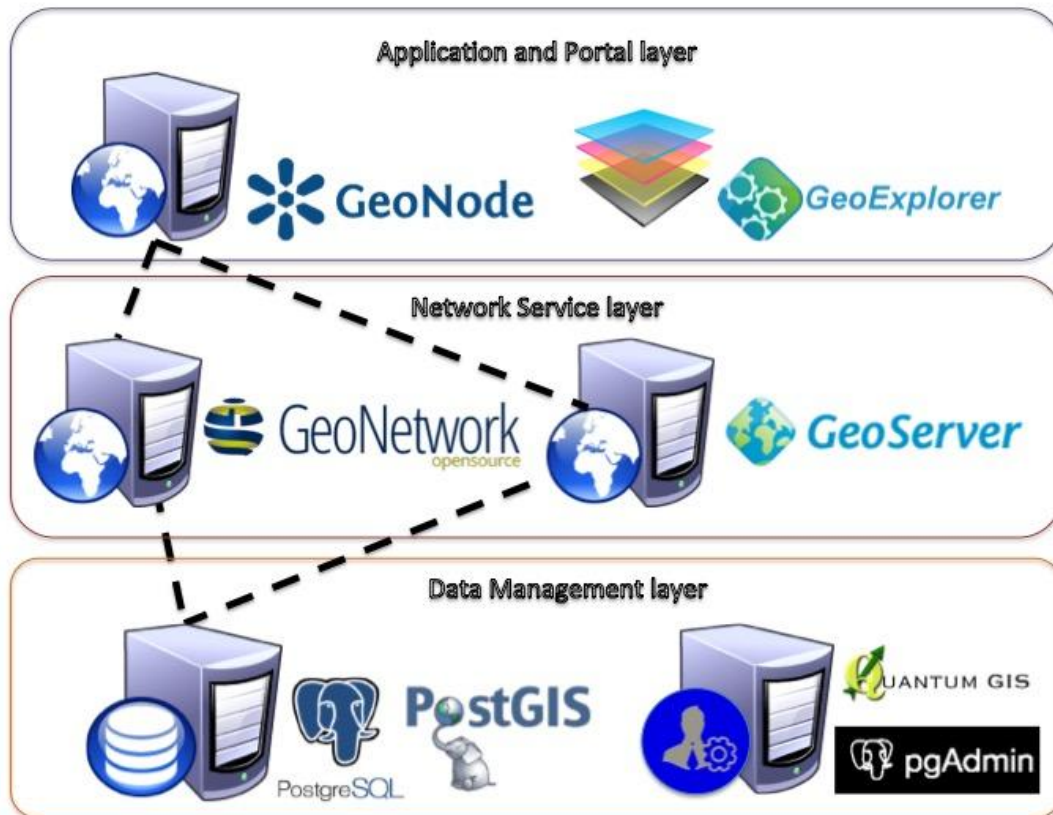
The software component to support the web service for data distribution and elaboration is based on the open source *Geoserver* that is interlinked with GeoNode and it is able to deploy web services for view (OGC WMS – View Map Service), for download and query select (OGC WFS – Web Feature Service, OGC WCS - Web Coverage Service) and for specific elaboration and calculation (OGC WPS – Web Processing Service).

### Portal and Client system

The *InfoMapNode* is the platform that represents the portal for the management and publication of geospatial data in a collaborative way, but also for developing geospatial information systems (GIS) that is based on the software package *GeoNode*, where most of the other components are embedded in order to simplify to the user the possibility to create and manage data, layer, map and metadata repository. In fact Data management tools built into GeoNode have the possibility to provide to user an integrated creation of data, metadata, and

*Implementation of the Shared Environmental Information System (SEIS) principles and practices in the ENP South region – SEIS Support Mechanism (ENI SEIS II South)*

map visualizations. Each dataset in the system can be shared publicly or restricted with different rank of authentication, in order to allow access to specific users only. In the platform, it is also integrated the client part based on *GeoExplorer* software component to allow to the user the view and publication of data and maps.



*Figure 5. INFO/RAC Spatial Data Infrastructure (SDI) to support ENI SEIS South II project.*



## **Software components to build the SDI**

In the previous section has been shown the architecture of Spatial Data Infrastructure developed by INFO/RAC for ENI SEIS South II project purposes. Here a more detailed technical description for each software component used to build up the SDI will be described. In a specific chapter will be introduced the user guideline to use each one within the InfoMapNode.

### PostgreSQL and PostGIS

PostgreSQL is a powerful, open source object-relational database system (ORDBMS) that uses and extends the SQL language combined with many features (complex queries, foreign keys, updatable views, multi-version concurrency control, data type, etc.) that safely store and scale the most complicated data workloads.

PostgreSQL has a proven architecture, reliability, data integrity, robust feature set, extensibility, moreover it's runs on all major operating systems, has been ACID-compliant since 2001, and has powerful add-ons such as the popular PostGIS geospatial database extender.

PostgreSQL is extensible software where is possible to custom functions, even write code from different programming languages without recompiling your database.

PostGIS is an open source extension for the PostgreSQL Database Management System that adds spatial feature functions such as distance, area, union, intersection, and specialty geometry data types to the database.

For the installation, configuration and SQL syntax to use to manage the data we redirect to official documentation of Postgres, available at: <https://www.postgresql.org/> and PostGIS extension: <https://postgis.net/>.

The INFO/RAC Infrastructure has built on PostgreSQL version 9.5 and PostGIS 2.4.3; this software as open source component can be replicated in the project countries, using the same approach and the same license of distribution.

### pgAdmin

pgAdmin is a graphic front-end software developed to support Postgres database in the management agent and it's mainly used to view, insert and update the data in the tables. With pgAdmin we can create a database instance from the scratch, defining the ownership and the functionality to be available for each user. It supports all PostgreSQL features, from writing simple SQL queries to developing complex databases or to query an active database (in real-time), allowing to the user to control modifications and implementations.

In the SDI has been installed the version 3.0 and also in this case for use guideline we resend to the official web site and documentation available at: <https://www.pgadmin.org/>.

### QGIS

This software is GIS open source desktop software, mainly used to view and manage the geographical information and data, it supports a number of raster and vector data formats, with new format support easily added using the plugin architecture. From its functionality is available the *DB Manager Plugin* that is one of the main and native tools to integrate and manage spatial database formats supported by QGIS (i.e. PostGIS, SpatiaLite, GeoPackage, Oracle Spatial, MSSQL, DB2, Virtual layers) in one user interface. In the SDI suite we use this

software to manage in the local domain geographic data but also to manage the PostGIS functionality within the database.

We can use different version and from the version 2.14 till the last these functionality are available and work fine; in the suite we use version 2.18 and for installation and user guideline we send to official web side <https://qgis.org>.

### Geonetwork

The GeoNetwork project started out in year 2001 as a Spatial Data Catalogue System for the Food and Agriculture organization of the United Nations (FAO), the United Nations World Food Programme (WFP) and the United Nations Environmental Programme (UNEP).

The software is widely used as the basis of Spatial Data Infrastructures all around the world, it's part of the Open Source Geospatial Foundation (OSGeo) and installation, configuration and use guide information can be reached at GeoNetwork website .

The Geonetwork software provides an easy to use web interface to search geospatial data across multiple catalogues. The search provides full-text search as well as faceted search on keywords, resource types, organizations, scale, etc. Moreover the users can easily refine the search and quickly get to the records of interest, GeoSpatial layers, but also services, maps or even not geographic datasets can be described as metadata in the catalogue.

Integrated in the suite with the version 3.2.2, it represents the main engine behind the SDI, is able to use different metadata template (ISO, Dublin Core, etc.), in order to collect information about the dataset and resource available in each country or in each organisation involved.

### Geoserver

The Geoserver is an open source software server written in Java that can use to share and edit geospatial data. It represents a web service deployment for interoperability, it publishes data from any major spatial data source using open standards based on Open Geospatial Consortium (OGC). It can read data from over a multiple spatial data sources and can produce output in many different formats to make spatial data more accessible to all.

GeoServer has a browser-based web administration interface application used to configure all aspects of GeoServer, from adding and publishing data to changing service settings. In the suite we have sutup the version 2.10, but more information regarding the installation and configuration as well as the possibility to download and use the most recent version are available at Geoserver official website: <http://geoserver.org/>.

### GeoNode and GeoExplorer

The GeoNode is platform for the management and publication of geospatial data. It brings together mature and stable open-source software projects under a consistent and easy-to-use interface allowing non-specialized users to share data and create interactive maps. The software is based on core components which are interoperable and as such, it is straightforward for you to integrate with external applications and services. The GeoNode platform (version 2.6.1) includes as default software to setup the Spatial Data management system (but can be integrated with other software components installed and configured separately) the following are the internal software elements:

- *Django. (It's the web framework written in Python and represents the panel of the content management system to administrate the platform based on GeoNode).*
- *psCSW (ISO metadata editor and catalog with limited functionality)*
- *GeoServer (Web service deployment software)*

- *PostgreSQL and PostGIS (Database and spatial extender plug-in)*

Moreover it's tool for integrated creation of data, metadata, and map visualizations in one place. From this interface, one can view and modify existing spatial layers and maps, as well as find information on other GeoNode users. It has the ability to restrict the access on your layers, maps, and documents to other users or group of users.

It is also designed to be a flexible platform that software developers can extend, modify or integrate against to meet requirements in their own applications.

### ***Environment indicators Data flow mechanism***

In the ENI SEIS South II project, the countries have agreed a set of H2020/NAP indicators on water, industrial emission and solid waste including marine litter. In order to help the collecting data process, in parallel with the Spatial Data Infrastructure that is prone to give and share data, layer and map, INFO/RAC has developed a specific module in its Data Centre.

The Data Centre is set of modules that allows INFO/RAC to acquire data for reporting action with specific quality check control, in order to collect data in standard and consistent way.

In the Data exchange from countries to international level, we need to define in the Data Centre the schema of data to acquire based on a standard template that should be defined in agreement with the parties. The structure of data to be delivered is defined as a dataset, which is a collection of tables that represents the reported data. Usually they use different exchange formats (CSV, XML, GML, etc.) coming from a MS Excel files, small databases or unstructured data documents.

The possibility to share data using a specific dataset schema defined as Data Dictionary in the Data Centre allows to the system to make a series of formal data model and content controls; the consistency of these controls represents the quality check mechanism.

#### *Data Dictionary description*

To define the structure of data flow or a dataset we need to define and agree on the Data Dictionary (DD) that is the definitions of datasets, tables and data elements exchanged; without DD it would be not possible to build the data model, which is behind the data flow and that is the basic element to build input data template to acquire the reported data.

They are defined by a set of attributes, the core set of which corresponds to ISO 11179 standard for describing data elements. The whole attribute set is flexible and Data Centre administrator can dynamically adjusts attributes from/to the system. Each Dataset consists of tables and each table is defined by its columns called data elements in the DD.

The data elements are built as blocks of tables and in DD data elements are divided in two groups:

- non-common elements belong into a certain single table and they cannot be used in any other table. They are defined within that certain table and they do not have a “life” outside of that table.
- common elements are not strictly attached to a single table. Instead, they can be used in many tables. They are potential candidates to be used in many datasets and tables and thus have been harmonised to avoid repetitions. A perfect example of a common element is CountryCode, which can modeled as a vocabulary list.

Each data element can be divided by different value domain in:

*Implementation of the Shared Environmental Information System (SEIS) principles and practices in the ENP South region – SEIS Support Mechanism (ENI SEIS II South)*

- Data elements with fixed values (i.e. code lists terms).
- Data elements with quantitative values (i.e. measured values), for these elements DD enables to define suggested values. The latter are DD's suggestions for data suppliers who are not sure which values exactly to supply.
- Data elements with values from a vocabulary have a linked vocabulary. The vocabulary concept values are valid values for the common data element.

Datasets, tables and elements in DD are defined by a set of attributes, the core set of which corresponds to ISO 11179 standard for describing data elements. The rest of the attributes are usually specific to Reporting rules and in any case the attribute set is flexible. Most common attribute of all is 'Name', standing for the name of the defined object.

In ISO 11179 two types of attributes are recognized and used in DD as well:

#### Simple attribute

Every such attribute is a name-value pair. They represent the same concept of attributes as in ISO 11179 part 3. These attributes are sometimes called repeating attributes. A good example of such an attribute is 'Keywords'. It is natural that there can be many keywords describing an object.

Important simple attributes when defining a data element in DD are:

#### Datatype

Indicates if the element's values are strings, integers, booleans, etc. Also specifies if decimals are to be used, and how many decimals are accepted.

#### MaxSize/MinSize

The max/min length of a data element value in terms of digits or characters. For example if an element's MaxSize is 4, its value '1,000.0' is one decimal too long. Or if an element's MinSize is 5, its value '603' is one digit too short.

#### MaxValue/MinValue

Can be applied only to data elements of numeric Datatype and they limit the element's numeric value. i.e. value '9.55' is too much if MaxValue is '9'.

#### Complex attribute

A complex attribute is almost the same as a simple attribute. The difference is that while a simple attribute is a name-value pair, a complex attribute is a set of name-value pairs. Every such pair is called the complex attribute's field. For example address would consist of country, city and street.

Important complex attributes when defining a dataset, table or element are:

#### SubmitOrganisation

The organisation that has added or changed the data element definition. This is mandatory to specify for all dataset, table and element definitions. Standard fields of this complex attribute are: name, contact, email, url, address, etc.).

#### RegistrationAuthority

Institution authorised to register the data definitions. Fields of this complex attribute are: name, acronym, address, contact-Person, url, etc.

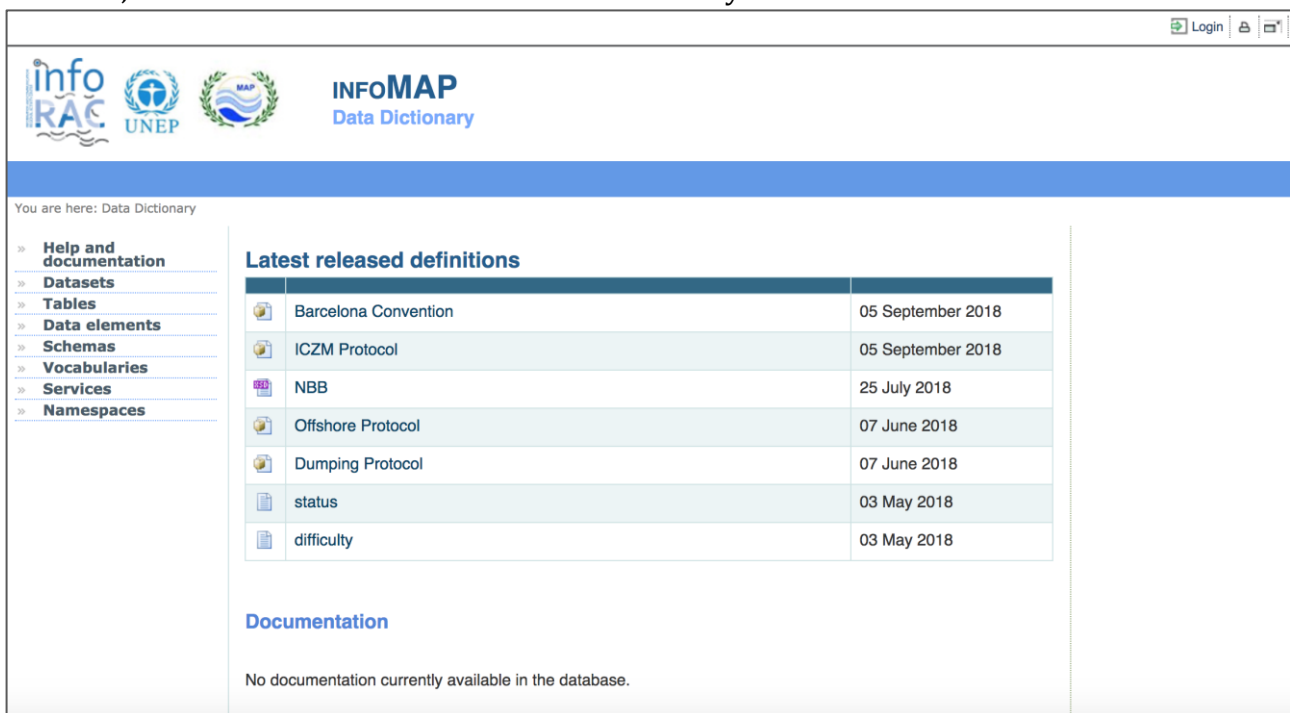
More about each and every attribute you can understand through using the DD, because each and every attribute in the DD is described in the specific document.

## Data Centre User guide:

To support the users to prepare the Report's submission or a specific dataset procedure we need that the Data Dictionary has been defined and implemented in the Data Centre (DC). The DD cycle starts with the factsheet that define in general way a purpose of data collection, passes to the Tables DD definition that describe in detail the content of each table that composes a dataset and goes to DD Schema definition that is the technical document where the tables are transformed in a data model and a specific schema (xsd) shall implemented in system in order to allow data acquisition and data quality check.

Within the DC there are 3 parts that involve the country in the report production: DD Library, Data Repository and WebForm for reporting.

The DD Library is schema repository, it's also a open section (figure 6) navigable of INFO/RAC Data Centre, where the user can navigate to view Datasets flow, Tables, Data Elements, Schemas, Vocabularies and Services available in the system.



Definition	Release Date
Barcelona Convention	05 September 2018
ICZM Protocol	05 September 2018
NBB	25 July 2018
Offshore Protocol	07 June 2018
Dumping Protocol	07 June 2018
status	03 May 2018
difficulty	03 May 2018

Figure 6. Data Dictionary section in the InfoMap Data Centre.

The Data Repository is web page where each country has a directory open only to specific user (need a reporter account to enter in), where the report files will be stored to check and view.

The Web Form for reporting is a page available when you enter in the own data repository, from this you can fill-in the form to produce the report or indicator dataset directly on web, or if the country is able to provide data flow in standard format following the DD schema or Template, the user is able to charge the XML file or a spread sheet template in CSV that generate automatically the data flow and make all the quality check to be compliant with DD Specific schema.

The Reporting or data flow cycle will be closed after some quality assessment control made by supervisor organization. In the case of ENI SEIS these will be: **EEA, UN Environment Programme/MAP components**; the final dataset stored as xml file in the Data repository will be upload in a specific central database managed by INFO/RAC. .

The aggregate data and assessment maps or documents will be deployed in the *InfoMapNode*



portal for public consultation and view.

In a separate document “SDI and Data flow User Guide” will be illustrated how the country user can manage and enter in the *InfoMapNode geoportal* to upload and create data and layer and how a country can provide a dataset for reporting system of H2020 Indicators within the ENI SEIS South II project.

### **References documentation for use software components**

PostgreSQL 9.5.14 Documentation: <https://www.postgresql.org/docs/9.5/static/index.html>

PostGIS 2.4.5dev Manual: <http://postgis.net/stuff/postgis-2.4.pdf>

pgAdmin III user guide: <https://www.pgadmin.org/docs/pgadmin3/1.22/>

QGIS 2.18 User guide: <https://docs.qgis.org/2.18/pdf/en/QGIS-2.18-UserGuide-en.pdf>

GeoNetwork 3.2.2 a guide for people who will be using the catalog web interface: <https://geonetwork-opensource.org/manuals/3.2.2/eng/users/html/overview/index.html>

GeoServer 2.10.x User Manual: <https://geoserver-pdf.readthedocs.io/en/latest/index.html>

GeoNode user workshop: <http://docs.geonode.org/en/latest/tutorials/users/index.html>