

Geospatial Data Infrastructure Components Deployed for LTER -Europe Researchers' Community

T. Kliment^{1,3}, P. Carrara², A. Oggioni²

¹ Faculty of Geodesy, University of Zagreb, Croatia

² Institute for Electromagnetic Sensing of the Environment, National Research Council of Italy, Milan, Italy

³ Horticulture and Landscape Engineering Faculty, Slovak University of Agriculture in Nitra, Nitra, Slovakia

Abstract

Assessing the status and trends of the environment requires the collection, management and publication of spatially referred observations. Since many years, Long-Term Ecological Research sites in Europe collect ecological data, resulting in long-term data series. Nowadays, advanced software tools can enable discovering, accessing, and distributing collected data in a user friendly way. Based on Open Geospatial Consortium (OGC) standards, Web based Geographic Information System allows access to interoperable distributed repositories of observations. The present paper gives an overview of the methods and solutions proposed and tested in the LIFE+ project EnvEurope for the community of researchers of the Long-Term Ecosystem Research (LTER) Network in Europe. These solutions consist in a straightforward online environment for metadata management and discovery, shaped on the ecological community and its practices; components of a Spatial Data Infrastructure including both a network of repositories deploying observations via OGC SOS (Sensor Observation Service) services and Web user interfaces to access and visualize them. The success of the presented approach is linked to the development and availability of easy to define, ready to use tools, enabling site managers to create their own repositories and services.

Keywords:

Geospatial data infrastructure (GDI); metadata; observations; LTER Network; DEIMS; INSPIRE; Sensor Observation Service (SOS).

Kliment, T., Carrara, P., Oggioni, A. (2015) "Geospatial Data Infrastructure Components Deployed for LTER-Europe Researchers' Community", *AGRIS on-line Papers in Economics and Informatics*, Vol. 7, No. 4, pp. 64 - 78, ISSN 1804-1930.

Introduction

More than 400 Long-Term Ecological Research (LTER) sites are disseminated throughout Europe and are precious sources of ecological data collected since many years (Adamescu et al, 2007; Mirtl, 2010). These sites are even more useful when studies on climate change and assessment of environmental trends (Metzger et al, 2010) at different scales are becoming a priority in the socio-economic and political agenda. Data collected at LTER sites refer to three main natural environmental categories, i.e., forest, sea and freshwater. This is however a source of environmental heterogeneity, besides the diversity of languages, national laws and country-based practices, etc. Researchers at each local site struggle with following a balance between the heritage of long local tradition

of data collection and the new requirements of sharing, harmonizing and comparing data at the European level. Information Technology tools seem not to improve the situation, since they are often perceived as a further burden to the daily activities of ecologists. Easily and safely storing data, distributing them following harmonized rules, either maintaining local data centres or feeding remote data collectors, are all actions that can be felt as an extraneous, unwelcome task.

In this context, the challenges for new, advanced IT solutions are:

- To offer ecologists user-friendly, online tools to enter and to store their own observations in a way which is respectful to both community practices and to the new European directive (INSPIRE

- Infrastructure for Spatial Information in Europe) (European Commission, 2007), which requires open data structures.
- To allow research sites both to maintain their own data repositories and to share data in an easy, interoperable way.
- To create user interfaces on the Web, allowing on the one hand discovering the provided data, on the other hand to subsequently visualize and access them, thus fostering data analysis.

The LIFE+ project EnvEurope (Kliment et al, 2013; Pugnetti et al, 2013) intended to initiate the creation of a common IT platform and to provide standards and guidelines for the entire LTER-Europe network (Mirtl, 2010). Within the project, a pilot platform for LTER-Europe was created in order to meet the above listed challenges, including:

- An online tool that allows entering metadata of sites, researchers and observations, linked together; the same tool allows online discovery of the information of interest filtered by location, time range, collected parameters, country, etc. Metadata follows the Ecological Metadata Language (EML) metadata specifications (Michener, 2006), but are also harmonized with the European directive INSPIRE metadata regulation (European Commission, 2008).
- Geospatial Data Infrastructure (GDI) components as Catalogue Service for Web (CSW) deploying compliant metadata catalogue, geospatial data services providing access to a network of distributed repositories of observations shared as OGC SOS services and online user interfaces to access and visualize sites locations, observations offered and their trends in time.

Materials and methods

1. Metadata online tool

In order to enable data discovery, interpretation and, if applicable, data analysis, information about the “how, where, when, what, who ...” needs to be captured in an accessible and understandable manner (Nogueras-Iso et al, 2005). If adequate metadata is available, data can be reused after years or decades, either on its own or in combination with data from other sources (Karasti et al, 2008). In the EnvEurope project, an online metadata tool was developed based on three metadata models in order to facilitate the LTER-Europe network

community of researchers in information discovery and sharing. The tool is based on the Drupal Ecological Information Management System (DEIMS) developed by the US LTER.

1.1. Metadata models

A complete model defining the descriptive information – metadata integrated in the system was defined through the conceptual schema overview including three metadata models as follows:

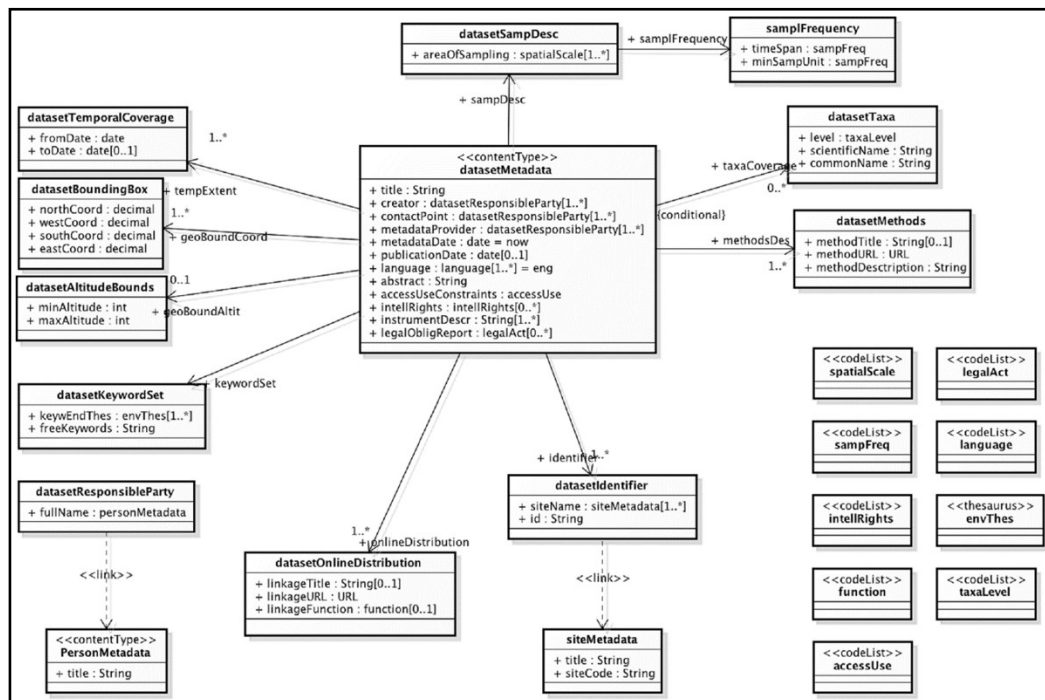
1. Research Site Metadata Model (RSMM) (Peterseil et al, 2012): To get a fast overview about the participating sites in a network there is a strong need for metadata about the inner structure, administration, data management and observations taken at the site or platform level. The previously used LTER InfoBase system was designed as a tri-component system by ALTER-Net providing site level metadata which described the LTER Sites and LTSE Platforms as a whole (Adamescu et al, 2010) and which formed the site network of ALTER-Net and LTER-Europe (Haberl et al, 2006; Mirtl et al, 2007). The RSMM application schema has been migrated from the LTER InfoBase into the DEIMS database and improved according to the user needs.
2. Person Metadata Model (PMM): The PMM serves several purposes. First, it provides a source for dataset metadata creators, contact points, and metadata providers. In addition, the person component serves as a personnel directory, allowing individuals to add information about their expertise, subjects investigated, role in their national LTER network, and networks in which they participate in addition to LTER-Europe. Persons can be associated with both datasets and publications. In order to be associated with a dataset, the person needs to be selected as a metadata creator, contact point, or metadata provider when creating a dataset metadata record. The PMM has been based on US-LTER model and extended for the user needs.
3. Dataset Metadata Model (DSMM) (Kliment et al, 2011) has been introduced for the first time into the LTER-Europe network community within EnvEurope. This brings a significant benefit for the community: researchers are able to share not only information about their research sites, or people who

perform individual tasks, but they can also bridge the former to information about the collected observational data instances aggregated as datasets. This information makes researchers easily aware of the existence of datasets from other research sites, which might be of interest for particular cross-domain analysis. The definition of metadata fields in the DSMM is based on the following categories of sources:

- a. Ecological Metadata Language (EML) - semantic description and syntactic definition of individual metadata elements was taken from the EML metadata specification (KNB, 2014), due to its recognition in the LTER networks around the globe.
- b. Global and European Environmental Data Infrastructures - Global and European data e-infrastructures define data themes with a close relation to the data collected within LTER-Europe network. For example, an infrastructure available through GEOSS divides information resources into nine social

benefit areas, out of which two are related to LTER domains - Ecosystems and Biodiversity. On the European level, infrastructure for geospatial information, which is defined by the INSPIRE directive has the main goal to provide harmonized geospatial data to support environmental and related policies in EU. Therefore apart of the reference data themes, several domain specific data themes as Species distribution, Habitats and biotopes, Environmental monitoring facilities, etc. have strong overlap with LTER.

- c. LTER-Europe network researchers expressed a set of requirements about the DSMM that were collected during the technical meetings. Beneficiaries were explained the meaning of dataset metadata and what would be the advantages and benefit of having their datasets documented by standardized information. The common metadata model was established and further implemented into the metadata editing form described later.



Source: own processing

Figure 1: UML class diagram modelling the dataset metadata class consisting of information resulted from requirements analysis.

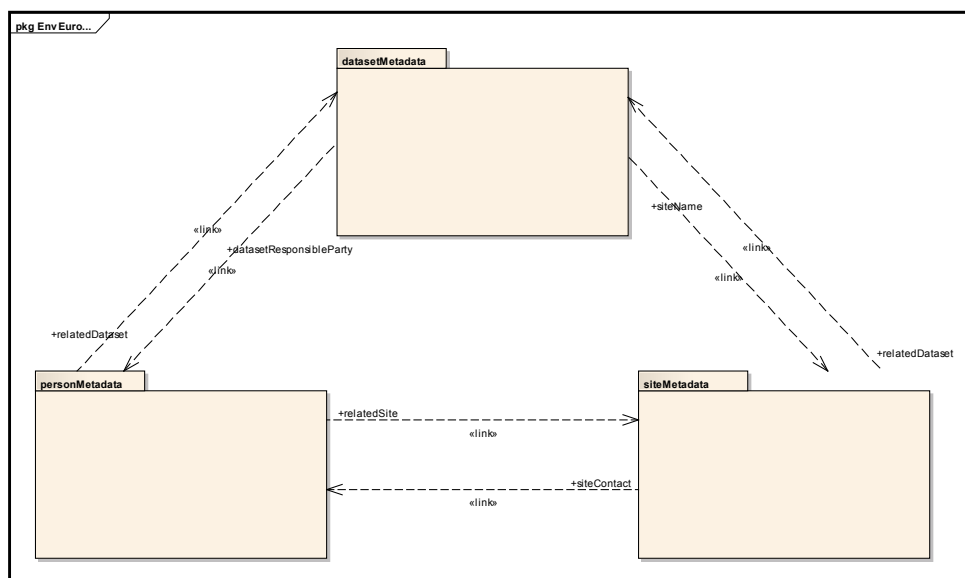
The main class `datasetMetadata` consists of 21 metadata categories that contain individual metadata elements of the DSMM (Figure 1). Detailed description of individual metadata categories and associated elements is provided in the metadata specification document (Kliment et al, 2011).

The UML package diagram in Figure 2 represents the relation of DSMM to other metadata models.

1.2. Implementation

Drupal Content Management System (CMS) was used to develop the metadata management platform as free and open source software package that allows easily organizing, managing and publishing content on the Web. This development environment has been chosen due to previously ongoing development actions carried out by the US LTER data management working groups. They developed first an ecological metadata editor based on Drupal CMS (San Gil et al, 2010) and continue in further improvements of the entire system. Drupal's code base is PHP scripting language and can be deployed in web servers, e.g. Apache HTTP Server. The underneath Database Management System (DBMS) used is MySQL; however, other DBMSs (e.g. PostgreSQL) are supported, too. JavaScript web scripting language is used to add additional functionalities for the data, modifying the content of HTML document displayed in the web browser, interacting with users and many other features. Individual pieces of data are represented as nodes, term well known in the Drupal community, which refers

to a record in a content table in the database. Each web page contains information from one or more nodes. Different types of information are represented by Content Types, which define groups of information for a specific category of information (e.g. Dataset metadata). Nodes of individual content types can be interlinked based on the conceptual definition and principles of relational databases. For example, one node from the Dataset content type may contain links to several nodes of the research site and person content types. Additional features as various field types, taxonomies, content views, search interfaces, etc. are available through a bulk of modules developed by the Drupal community of developers, and thus can be added to web pages without further programming works. Drupal Themes that define styles of the web page content displaying ensure the graphical user interface layout of the application front end. The web administrator can change some of the style characteristics (e.g. font size, font family, and grid layout). For more complex changes, a high-level knowledge of CSS mechanism is needed. Drupal provides an easy way to maintain the content of an information system within any area of interest. The tool required to interact with a Drupal site instance is an Internet browser, which almost all can perform the operations. In order to create the default Drupal site accommodating the requirements of a specific community (e.g. LTER-Europe), customization needs to be performed. By example, defining new content types, serving the content in various commonly used formats



Source: own processing

Figure 2: DEIMS metadata models designed with bidirectional links.

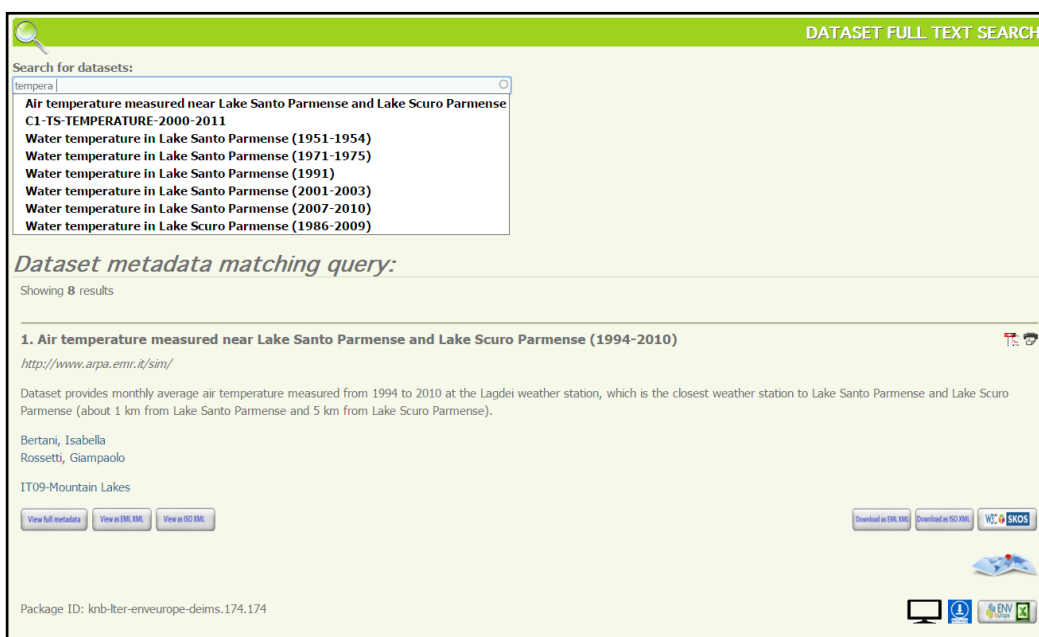
(XML, Excel spreadsheets, PDF files, etc.), or defining various views to offer the content in user friendly and intuitive ways. The Views module provides user-friendly definition interface for SQL queries together with the actual final web layout. Using the web GUI, authorized user with admin rights can configure the query (fields, filters), and the layout (style, page settings etc.) (Rumbauhg et al, 2004). Moreover, since its version 6, Drupal provides a module for taxonomy definitions in order to define semantic relationships among keywords within a particular area (e.g. LTER) ideally coming from a controlled vocabulary like EnvThes - The EnvEurope Thesaurus (Peterseil et al, 2012).

Based on the conceptual model described above, the dataset metadata content type was created in order to provide users with an online editing form. Individual categories (e.g. 19. Dataset sampling description) and nested metadata fields (e.g. Sampling time span) of various types (e.g. free text, autocomplete, select list, etc.) were defined. The editing form consists of three main sections: basic guidelines, editing section, and action buttons. The form has the following selected features that were developed to facilitate the metadata collection: i) linking to associated person and site metadata via autocomplete functionality; ii) Annotating datasets with concepts from EnvThes through autocomplete functionality with multiple value definition; iii) Geographic

extent definition with an automatic geographic bounds calculation performed from multiple polygons drawn on a map; iv) Instrumentation, Sampling description and Methods taken from the ECOPAR (Parameters and Methods for Ecosystem Research & Monitoring) database and provided through autocomplete functionality; v) List of predefined values with option to add new values - spatial scale, sampling time span, minimum sampling unit, intellectual rights, etc.

The metadata collected serve primarily for data discovery. Additionally, data evaluation can be performed in order to support the comparison of similar datasets. Therefore, several searching interfaces were developed combining functionalities of several Drupal modules. Namely, Views, Faceted Search, Finder, etc. were used. The following searching clients are available:

- Simple full text search: provided for each information type (dataset, person, and site) separately with simple searching box querying titles of metadata records available in the system extended by autocomplete functionality (Figure 3).
- Guided search: available for all information or each type individually. A searching pattern provides a keyword search with a possibility to refine the query with the predefined categories (resource type, organization, language, etc.).



Source: own processing

Figure 3: Simple full text search user interface for dataset metadata and result records with links to metadata and data representations.

2. Geospatial data infrastructure components for sharing LTER features and observations

Concepts like interoperability are fundamental in realizing a linkup among data characterized by spatial (e.g. depth, geographical projection or location, relative position), temporal (e.g. time zone) and thematic (e.g. quality, domain, unit of measurement) attributes. Having the necessity to exploit the interoperable access to observations from multiple sites, using heterogeneous sensors, issues to be faced are:

- Heterogeneity: LTER-Europe observational network is heterogeneous; managing authorities are numerous and have different skills, resources and IT expertise. Network nodes technologies are not homogeneous in the collection, frequency and distribution of the measured parameters (e.g. different temporal and spatial resolution, units, identifiers).
- Quality check and harmonization: To allow for a comparison between data coming from sites, quality check of data is a priority. Exclusion of outliers, comparisons among nearby stations, and trend analysis at different temporal granularity are operations that must be carried out at different levels of the data processing workflow; they would allow an effective and meaningful comparison. Another important action is the harmonization of collection and storing practices in order to improve the overall quality of the observations collected from the network.
- Description and history of sensors: Information on sensors used like their description, searching keywords, identification, classification, characterization of physical properties or electrical requirements, capability, contacts of manufacturer, owner or operator, input, output and components of the system, and especially history log to track any changes or calibration, must be collected and made available in order to assess their quality, capacity, features and to compare the sensors and data measured by them.

The pilot created in EnvEurope aimed at testing how distributed repositories and a decentralized data infrastructure can tackle the above issues. By deploying data from heterogeneous, asynchronous sensors connected to the Web, the pilot wishes to prove how OGC services are

able to interoperate and to let sharing, visualizing and accessing observations and measurements, without forcing data replication and harvesting towards a centralized caching data centre.

2.1. Sharing LTER observations

In fact, data management within LTER realm is focused on observations from sensors. Therefore, a framework of standards has been proposed and supported by OGC under the common umbrella of SWE (Botts et al, 2008), which includes: SWE Common Data Model, Sensor Model Language (SensorML), Sensor Event Service (SES), Sensor Planning Service (SPS), Sensor Observation Service (SOS) for observations collected by sensors.

In the EnvEurope pilot, the SOS service (Na et al., 2007) has been adopted and tested. SOS has been developed for discovering, binding and querying individual sensors or sensors platforms in real-time (RT), near real-time (NRT) or delay mode (DM) (Bermudez et al, 2009). With SOS, two more specifications work together: SensorML for describing characteristics and capability of the sensors, and Observations and Measurements (O&M) for encoding observations and measurements.

SOS specifies a standard Web service interface for requesting, filtering, and retrieving observations and sensors' information (see Table 1 for the available requests categorized into core, transactional and enhanced, respectively). This interface is the intermediary between a client and an observation repository or near real-time sensor channel. Clients can also access SOS to obtain metadata information that describes the associated sensors, platforms, procedures and other metadata related to observations.

| | |
|--------------------------|--------------------------|
| Core Operations | GetCapabilities |
| | DescribeSensor |
| | GetObservation |
| | RegisterSensor |
| Transactional Operations | InsertObservation |
| | GetObservationById |
| | GetResult |
| Enhanced Operations | GetFeatureOfInterest |
| | GetFeatureOfInterestTime |
| | DescribeFeatureType |
| | DescribeObservationType |
| | DescribeResultModel |

Source: modified by Bermudez et al, (2009). For more information about operation, descriptions see (Na et al, 2007).

Table 1: Requests carried out in SOS service divided by type.

The components of the pilot in EnvEurope are data repositories and OGC services that are both distributed, "... to ensure that geospatial data are stored, made available and maintained at the most appropriate level ..." (see Art. 6 – INSPIRE Directive 2007/2/EC) (European Commission, 2007). This aspect is very important in the LTER context, where institutions from all over Europe need to manage and keep the data collected from their equipment.

The EnvEurope pilot provides data collected by sensors, especially physical and chemical parameters listed in Table 2.

| Parameter full name | SOS offering | Unit |
|---------------------------|--------------|------|
| Air Temperature | AirTemp | °C |
| Water Temperature | WaterTemp | °C |
| Total Phosphorus | TP | µg/L |
| Total Nitrogen | TN | µg/L |
| Phosphate | P-PO4 | µg/L |
| Ammonia | N-NH4 | µg/L |
| Nitrates | N-NO3 | µg/L |
| Silica | SI | mg/L |
| Secchi depth Transparency | SDT | m |
| Salinity | Sal | PSU |
| Chlorophyll | Chla | µg/L |

Source: modified by Bermudez et al, (2009). For more information about operation, descriptions see (Na et al, 2007).

Table 2: List of parameters provided by observations collected in EnvEurope.

The observation flow among different components of the architecture is organized by the standard SOS interface. Observations collected by local, distributed sensors are stored in the respective local repositories by the SOS InsertObservation() request. Also the dialogue between application layer and service/data layer occurs through different standard requests, e.g. GetObservations(), GetFeatureOfInterest(), DescribeSensor(). In this way, a complete decoupling is obtained between the components that store, distribute and deploy observations and the applications that allow end users to search and access them on the Web: in fact, the interoperable dialogue is guaranteed by standard requests, not depending on whichever implementing environment is used. In the EnvEurope pilot, decoupling has been proven by testing observations' access and exploitation by both advanced tools (i.e. RStudio Server and Taverna) and a plain GIS-like user client GeoViewer created in the presentation layer.

2.2. From metadata to geospatial observations of datasets

Linking mechanism between the metadata of dataset and respective associated dataset deployed through available services for observation portrayal (Web Map Service – WMS) and download (SOS). If the metadata of dataset provides all necessary information, which can be used to construct a link, an icon is displayed within the discovery interface (Figure 3). The icon representation depends on a type of data service availability for the dataset described by metadata.

All datasets published by SOS have correspondent metadata stored in the DEIMS and are accessible for authorized users directly from the discovery client. The DEIMS generates a URL defining SOS GetObservation() request, which is posted to the SOS server. An example of SOS GetObservation() URL is constructed as follows:

```
http://sp7.irea.cnr.it/tomcat/envsos/sos?
SERVICE=SOS&
VERSION=1.0.0&
REQUEST=GetObservation&
OFFERING=AirTemp&
PROCEDURE=urn:lter:object:feature:Sensor
LTER_U_IT_009-Lagdei:AirTemp&
OBSERVEDPROPERTY=urn:lter:def:property:
OGC:1.0.30:AirTemp&
FEATUREOFINTEREST=LTER_EU_IT_009
-Lagdei&
EVENTTIME=1994-01-01T00:00:00/2010-12
-31T00:00:00&
RESPONSEFORMAT=text/xml;subtype="om
1.0.0"
```

Individual parameters of the URL are derived from corresponding metadata fields as follows:

- OFFERING – the value is taken from the dataset keyword part comparing provided keywords with the list of parameters offered by the service – e.g. AirTemp;
- PROCEDURE – the value is taken from the dataset identifier – site name as LTER-Europe site code and station code, than parameter value concatenated with the sampling frequency value – e.g. urn:lter:object:feature:Sensor:LTER_EU_IT_009-Lagdei:AirTempMonthly;

OBSERVEDPROPERTY – the parameter value from offering section is appended

- to the default property URN - urn:ogc:def:property:OGC:1.0.30:AirTemp;
- FEATUREOFINTEREST – the value consists of a LTER-Europe site code derived from the site name in the dataset identifier and the station code - e.g. LTER_EU_IT_009-Lagdei;
- EVENTTIME – the temporal filter is defined by the values provided in the dataset temporal extent metadata section (From and To dates) – e.g. 1994-01-01T00:00:00/2010-12-31T00:00:00.

The response to request represented by this URL is a collection of observations (dataset) encoded in O&M XML data model.

2.3. From metadata to geospatial features of research sites

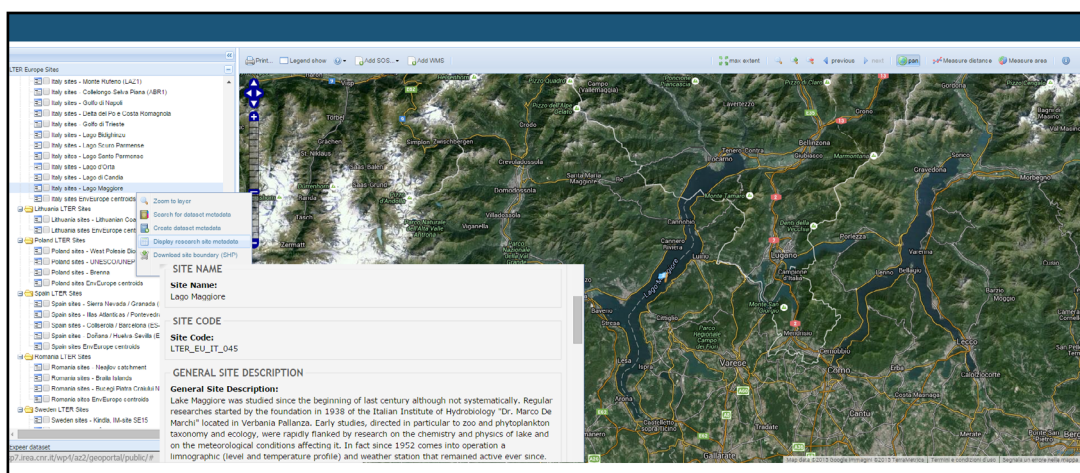
Geographical representation of datasets is implemented by linkage between DEIMS discovery client and GeoViewer using permalink with parameters latitude (lat) and longitude (lon). The values are defined as averages of the boundaries provided in the metadata section corresponding to the geographic coordinates bounding the dataset and calculated by the DEIMS. The objective of this functionality was to provide users an easy way to navigate from the descriptive representation to geographical location overview. A similar functionality is provided also for the research sites, where WMS GetMap() and WFS GetFeature() requests are constructed with values retrieved from the metadata fields Site Latitude, Site Longitude and Site Code. ESRI Shapefiles of EnvEurope sites were provided and made available to be portrayed by WMS together with the sites' basic information

in the GeoViewer (Figure 4). In addition, the sites' geospatial features can be downloaded as shapefile through WFS. On the contrary, reverse links from the site geographic portrayal in the GeoViewer to DEIMS components (metadata editor, dataset search, metadata preview) are provided, in order to drive the user from a site geospatial feature to its metadata as well as linked datasets' metadata.

3. Monitoring use cases implementation using deployed GDI components

The presentation layer of the pilot contains a client in the form of a GIS-like user interface, called the GeoViewer. It has been created to perform tests in the case when potential non-skilled users are operators involved in daily ecological tasks, such as marine monitoring. Its characteristics are briefly described here in terms of tasks and data, listing services coping with them and the solutions adopted in the pilot. The main testing actions they would perform to assess seawater quality through a network of heterogeneous, distributed stations of sensors may be summarized in the following three use cases:

1. To retrieve and display a description of the station, of sensors available, and of the measurement processes (e.g. calibration, gain, accuracy, offset, etc.) which could include quality control of all sensors of a station;
2. To retrieve and comparatively display observations of one selected parameter (e.g. water temperature, wind direction, wind speed, etc.) collected from multiple, distributed sensors;



Source: own processing

Figure 4: Research sites geospatial features' map representation and links to metadata and data download services.

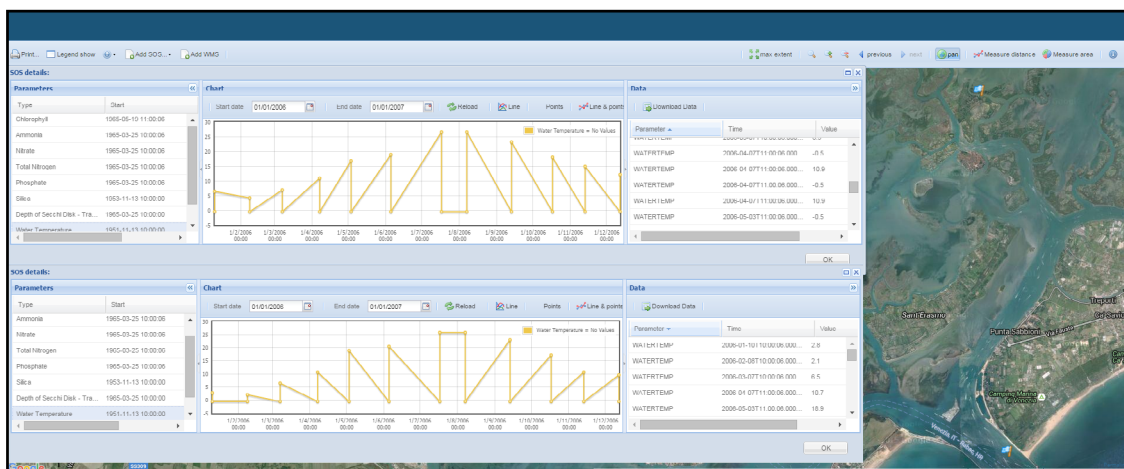
3. To retrieve observations of all parameters collected from all sensors from a specific station.

These use cases were tested by real examples as follows:

1. The user may want to know the features of the thermometers of marine stations in LTER. To display on a map the location of different water temperature sensors (thermometers), the pilot system exploits the Enhanced Operation GetFeatureOfInterest() (Table 1) to obtain the coordinates of stations participating in the pilot. The user's selection (by clicking) activates a GetCapabilities() request that produces the parameters each station collects. This way the user can select stations that provide data on water temperature; she/he can then select one of them and, after selection, the client performs a DescribeSensor() request. This allows to display the SensorML description of the water temperature sensor with general description, keywords, identification, classification, characterization of physical properties, electrical requirements, capability, contacts of manufacturer, owner or operator, input, output and components of the system, and moreover its history log to track any changes or calibration.
2. The user in this case may want to retrieve the water temperature in the whole North Adriatic Sea during summer (from 20 of June and 23 of September) 2011, and to know

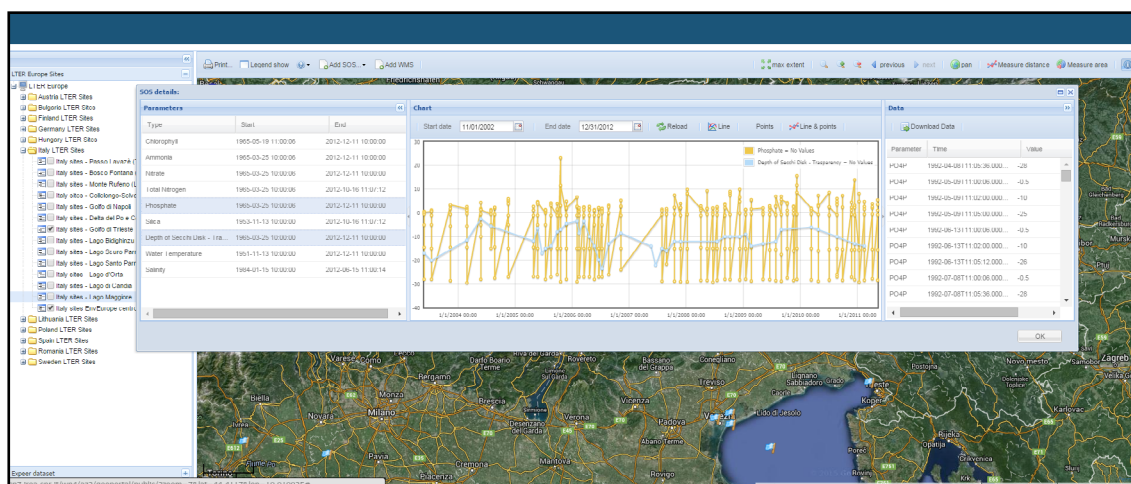
the geographic position of sensors. The pilot system performs a GetCapabilities() request, and shows stations providing data on water temperature and whose observations cover the period requested by the user. In fact, the response to this request contains, among other, information about: parameters measured in each station, period covered by different sensors, and geographic position. The second step is to make graphical representations of the observations through charts. The request GetObservation() with time period filtering can be used to get observations from all stations that satisfy the filter. An example of the results can be seen in the Figure 5.

3. The user in this case may want to retrieve all data collected by all sensors in one station (Figure 6). To this aim the pilot system simply exploits GetCapabilities() and GetObservation() requests to list the parameters and the corresponding values, respectively. The SOS that serves observations from the station selected by the user on the client map can be queried independently and it lists all observed properties present in the response capabilities.



Source: own processing

Figure 5: GeoViewer is able to visualize information distributed by multiple OGC SOSs, such as the related stations' positions and graphic visualizations of parameters trend (water temperature) during a given period.



Source: own processing

Figure 6: GeoViewer is able to comparatively visualize trends of many parameters distributed by an OGC SOS (here phosphate and depth of Secchi's disk) during a given period.

Results and discussion

DEIMS is able to provide a common portal to describe and access the LTER-Europe information resources in terms of research sites, researchers and dataset; DEIMS is an easy tool to collect and discover metadata of these resources. Though DSMM aims to cover all three levels of metadata usage (discovery, evaluation, and synthesis), its primary focus is devoted to the discovery and the evaluation levels. In other words, it focuses on those content areas that will allow the discovery of datasets that would potentially be useful for meta-analysis or other synthetic activities; and to perform a preliminary evaluation of the suitability of the dataset for further analysis or synthesis. Dataset metadata allow to carry out the further step of data download and exchange process, thus to evaluate the fitness for use of the data and finally retrieve information on how to use the data and, if available, how to display, how to access, how to transform, or process, etc. The testing phase of the pilot has brought significant value of the information, collected by means of the DEIMS GUI (Graphic User Interface) available for the researchers, e.g. more than 400 metadata records describing datasets, thus set of observations. Observations are deployed on the Web exploiting an SOS service; they are aggregated into individual datasets described by metadata. The result consists in more than 216 thousand of individual observation instances with a time coverage of more than 60 years (the first observation instance dates to 15/01/1951 and the latest to 16/10/2012) and geospatial coverage in the scope of 27 stations collecting

observations within 21 LTER-Europe sites from eight European countries. Additionally, all the observations collected by researchers as aggregated datasets are available to be displayed and used for further analysis from the EnvEurope Data Server through SPARQL queries invoked directly from the DEIMS discovery client, i.e. transparent to the user. Regarding the descriptive information about research sites and individual researchers, 442 metadata records are available for the LTER -Europe research sites and 862 records describing individual researchers. Dataset metadata can be downloaded and exchanged as EML, ISO19139 or RDF files. Another result obtained in the pilot is the GeoViewer designed and implemented in order to make collected observations understandable and visible in a simple way for researchers as an easy web-based interface to provide further ways of representations, e.g. geographical or statistical. It has been tested by means of three use cases depending on the researchers' requirements. Based on the tests performed by researchers, the evaluation of 2 out of 3 proposed use cases can be considered as satisfactory, thus facilitating the research work.

In general, the architecture, components and implementation solutions proposed in the pilot revealed to be able to cope with the requirements of a community of ecology researchers wishing to retrieve and display observations coming from heterogeneous sensors on distributed stations, stored in distributed repositories connected to the Web and delivered via standard OGC Web services in the SWE framework. The multilayer structure and the service approach enable decoupling

of components; in particular, each Institution hosting and maintaining a sensor station is allowed to store observations and to deliver them to multiple independent clients, in a standard, interoperable way, well recognized and accepted at European and global scale. Quality check and harmonization are fostered by the multi-layered approach that allows to include components and tools aimed at those purposes at different level; by example a fast-track quality control can be performed before the storage of observations in the repositories, while a spatially extended cross validation process can be included in the application layer, where values from multiple sites are available.

If we consider the challenges defined in methodology section, the proposed GDI is able to cope with technological heterogeneity of the sites and sensors since it is based on the use of OGC standards, able to describe sites and sensors characteristics but offering a uniform way to communicate among the implementation components. Uniform metadata and shared sensor/observation models are also a way to describe, search and compare quality. However, they are even more useful in facing the need to provide descriptions of sensors and their status, information necessary e.g. to maintain the network and to compare the sensors' performance.

The development of data management and exchange systems, which is being performed in the US LTER network, has chosen a different approach (San Gil et al, 2009). This could have been influenced by several factors: (i) Different approach and understanding in the conceptualization process – metadata models, or content categories developed and being used in the US LTER are simpler, especially the research site model is significantly simplified comparing to the European one; but on the other hand further metadata models are included in the US infrastructure (e.g. Publications, Projects, Variables); (ii) Data access services - data are included in EML encoding, whereas in Europe only metadata are encoded in EML, while the data access services proposed in the pilot is OGC SOS; this results from different legislation requirements as well as ongoing data initiatives (Open Linked Data). The same situation appears also for metadata encoding and services provided by the infrastructure. While the US approach adopts a combination of EML with Metacat as a catalogue, the European approach has taken into account requirements defined by INSPIRE and SEIS and therefore developed a mapping of EML metadata in the ISO 19115/19139 schema encoding, which is the application schema for the INSPIRE metadata

(Kliment et al, 2012). Additionally, in order to provide an INSPIRE Discovery Service interface, a catalogue service (CSW) by has been deployed, configured and populated by the metadata collected via DEIMS. Therefore, metadata of dataset collected in the European pilot are compliant with both EML encoding (to allow comparison and integration with the US LTER community) and INSPIRE. It is worth noticing that the dataset included in the pilot are mainly related to abiotic parameters and their description. The inclusion and description of biotic measurements require a further effort that has only been envisaged in EnvEurope.

Conclusion

The current global environmental research scene is highly fragmented, by disciplines or by domains, from oceanography, life sciences and health, to agriculture, space and climate. When it comes to cross-disciplinary activities, the notions of "building blocks" of common data infrastructures and building specific "data bridges" are becoming accepted metaphors for approaching the data complexity and enable data sharing. Data originating from a huge number of research projects, just completed or on-going, realized within the research communities, are becoming more visible due to several positive factors: (i) researchers have understood that publishing their data on the web in de-facto standardized way brings significant added value to their daily work; (ii) Legislation driven initiatives supporting the development of data infrastructures in the public sector, e.g. PSI, INSPIRE, SEIS, GMES have influenced research sector in a positive way, which is resulting in several initiatives and projects in ongoing (e.g. Research Data Alliance) or a proposal phase (e.g. Refer, or Biounify Cost (Co-operation in the field of Scientific and Technical Research) action project proposals) (iii) Links of large datasets from scientific research data in the relationship between Big data and Open data. Big, open data does not come always from governments: More and more scientists are sharing their research in astronomy, genomics, and other areas in a new, collaborative research model. Other researchers are using big data collected from social media – most of which is open to the public – to analyse public opinion and market trends (Gurin, 2014).

This paper tried to offer a contribution to this current debate, by describing the activity performed for the creation of the metadata tool and a pilot component of GDI, exploiting OGC SOS services,

in the framework of LTER-Europe ecology network. DEIMS and the pilot have been implemented and tested within the EnvEurope project; the results obtained are also described in the paper. In particular, DEIMS and the pilot approach (mainly the GeoViewer user interface) have been positively judged in other projects and initiatives dealing with heterogeneous environmental observations such as RITMARE, LifeWatch, or NextData.

The authors do not hide that the job to be done is great: in particular, the technological development of the tools to implement SWE components (and in particular SOS) is still overwhelming for the community of researchers in ecology. The success of the approach is linked to the development and availability of easy to define, ready to use tools, enabling site managers to friendly create their own repositories and services. Cloud providers can also offer a solution to the security issues raised by service distribution in small institutions. The beauty of this user friendly web geoservice-based data management tool, is the possibility of a wide usage and application to different sources; for instance in the determination of abiotic parameters like wind erosion intensity using soil particles (Lackóová et al., 2013), where the data collected in the field can be used as components/ input to create a repositories to be shared

with the community. Another perspective is related to syntactic and semantic harmonization of metadata and dataset, which requires intelligent applications that integrate the current technological solutions and standards with knowledge coming from the domain experts. Additionally, the conformance and validation of individual components of the pilot described in the paper will need to be tested against the requirements defined by legislation and related standardization as described in several researches works (Cibulka, 2013; Horák et al., 2011; Kliment et al, 2012; Lopez-Pellicer et al., 2014). The goal is to ensure an appropriate level of interoperability and thus bridge pilot's components with other relevant European and worldwide information infrastructures.

Acknowledgements

The work reported in this paper was funded by the EnvEurope © 2010-2014 Life Environment Project LIFE08 ENV/IT/000399 (<http://www.enveurope.eu/>), and the LifeWatch project (<http://www.lifewatch.eu/>). The authors would like to thank the editors and reviewers who have given so generously of their time to assess the paper. The authors are grateful for their advices in dealing with the manuscripts.

Corresponding author:

Ing. Tomáš Kliment, PhD.

Faculty of Geodesy, University of Zagreb,

Kačićeva 26, Zagreb, Croatia

Phone: + 385 (1) 4639 367, E-mail: tomas.kliment@geof.hr

References

- [1] Fadamescu, M., Cazacu, C., Peterseil, J., Datcu, S., Schleidt, K. Report on LTER InfoBase, 2007.
- [2] Adamescu, M., Peterseil, J., Dactu, S., Cazacu, C., Vadineanu, A. Elements for the design of a General Ecological Database. In: Maurer, I. and Tochtermann, K. (eds.) Information and Communication Technologies for Biodiversity and Agriculture. Shaker Verlag, Aachen, 2010. p. 49-66. ISBN 978-3832284596.
- [3] Bermudez, L., Cook, T., Forrest, D., Bogden, P., Galvarino, C., Bridger, E., Creager, G., Graybeal, J. Web feature service (WFS) and sensor observation service (SOS) comparison to publish time series data. In: Collaborative Technologies and Systems, 2009. CTS'09. International Symposium on. IEEE, 2009. p. 36-43. ISBN 978-1-4244-4584-4.
- [4] Botts, M., Percivall, G., Reed, C., Davidson, J. OGC® sensor web enablement: Overview and high level architecture. In: GeoSensor networks. Springer Berlin Heidelberg, 2008. p. 175-190. ISBN 978-3-540-79995-5 (Print), 978-3-540-79996-2 (Online).
- [5] Cibulka, D. Performance Testing of Web Map Services in three Dimensions—X, Y, Scale. Slovak Journal of Civil Engineering. 2013, Vol. 21, No. 1, p. 31-36. e-ISSN 1210-3896, ISSN 1338-3973.

- [6] European Commission (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Published in the official Journal on the 25th April, 2007.
- [7] European Commission (2008). Commission Regulation (EC) No. 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards Metadata.
- [8] Gurin, J. Open data now: the secret to hot startups, smart investing, savvy marketing, and fast innovation. McGraw Hill Education, 2014. ISBN 978-0-07-182977-9, e-ISBN 978-0-07-182978-6.
- [9] Haberl, H., Winiwarter, V., Anderson, K., P., Ayres, R. U. From LTER to LTSER: Conceptualizing the socioeconomic dimension of long-term socioecological research. *Ecology and society*. 2006, Vol. 11, No. 2, p. 256-289. ISSN: 1708-3087.
- [10] Horák, J., Ardielli, J., Růžička, J. Performance testing of web map services. In: *New Challenges for Intelligent Information and Database Systems*. Springer Berlin Heidelberg. 2011, p. 257-266. ISBN 978-3-642-19952-3 (Print), e-ISBN 978-3-642-19953-0.
- [11] Karasti, H., Baker, K. S. Digital data practices and the long term ecological research program growing global. *International Journal of Digital Curation*. 2008, Vol. 3, No. 2, p. 42-58. ISSN 1746-8256.
- [12] Kliment, T., Oggioni, A. EnvEurope (LTER Europe) Metadata Specification for Dataset Level. EnvEurope Project Deliverable A1.1.2b, Project report, 2011. 82 pp.
- [13] Kliment, T., Oggioni, A. Crosswalk between EnvEurope (EML) and INSPIRE (EN ISO) metadata profiles – a step towards SEIS. Joint International Conference “Biological reactions of forests to climate change and air pollution”, 21-24 May, 2012, Kaunas, Lithuania.
- [14] Kliment, T., Peterseil, J., Oggioni, A., Pugnetti, A., Blankman, D. Life+ EnvEurope DEIMS -improving access to long-term ecosystem monitoring data in Europe. In: *EGU General Assembly Conference Abstracts*. 2013, p. 4920.
- [15] Kliment, T., Tuchyňa, M., Kliment, M. Methodology for conformance testing of spatial data infrastructure components including an example of its implementation in Slovakia. *Slovak Journal of Civil Engineering*, 2012, Vol. 20, No. 1, p. 10-20. e-ISSN 1210-3896, ISSN 1338-3973.
- [16] Lackóová, L., Halászová, K., Kliment, M., Urban, T. Wind erosion intensity determination using soil particle catcher devices. *Journal of Central European Agriculture*. 2013, Vol. 14, No. 4., p. 1347-1355, ISSN 1332-9049.
- [17] Lopez-Pellicer, F., J., Latre, M. Á., Nogueras-Iso, J., Zaragoza-Soria, J., Barrera, J. Behaviour-driven development applied to the conformance testing of INSPIRE Web services. In: *Connecting a Digital Europe Through Location and Place*. Springer International Publishing. 2014, p. 325-339. ISBN 978-3-319-03610-6, e-ISBN978-3-319-03611-3.
- [18] Metzger, M. J., Bunce, R.G.H., Eupen, M., Mirtl, M. An assessment of long term ecosystem research activities across European socio-ecological gradients. *Journal of environmental management*. 2010, Vol. 91, No. 6, p. 1357-1365. ISSN 0301-4797.
- [19] Michener, W., K. Meta-information concepts for ecological data management. *Ecological informatics*. 2006, Vol. 1, No. 1, p. 3-7. ISSN 1574-9541.
- [20] Mirtl, M. Introducing the next generation of ecosystem research in Europe: LTER-Europe's multi-functional and multi-scale approach. In: *Long-Term Ecological Research*. Springer Netherlands. 2010, p. 75-93. ISBN 978-90-481-8782-9.
- [21] Mirtl, M., Krauze, K. Developing a new strategy for environmental research, monitoring and management: The European Long-Term Ecological Research Network's (LTER-Europe) role and perspectives. *Nature conservation management—From idea to practical results*. 2007, Vol. 36-52, ISBN 978-8387414986.

- [22] Na, A., Priest, M. Sensor observation service. Implementation Standard OGC, 2007.
- [23] Nogueras-Iso, J., Zarazaga-Soria, F. J., Muro-Medrano, P., R. Geographic information metadata for spatial data infrastructures. Resources, Interoperability and Information Retrieval, 2005. ISBN 978-3-540-24464-6. e-ISBN 978-3-540-27508-4.
- [24] Peterseil, J., Kliment, T., Oggioni, A., Schentz, H., Blankman, D., Schleidt, K. Manual on Data Management Standards for LTER Europe. EnvEurope Project Report A1.4.1, 2012. p. 91.
- [25] Peterseil, J., Mirtl, M. Site description metadata schemata. EnvEurope Project Report A1.1.2a, 2012, p. 53.
- [26] A. Pugnetti, F. Acri, F. Bernardi Aubry, E. Camatti, E. Cecere, C. Facca, P. Franzoi, E. Keppel, A. Lugliè, M. Mistri, C. Munari, B. M. Padedda, A. Petrocelli, F. Pranovi, S. Pulina, C. T. Satta, N. Sechi, A. Sfriso, M. Sigovini, D. Tagliapietra, P. Torricelli. The Italian Long-Term Ecosystem Research (LTER-Italy) network: results, opportunities, and challenges for coastal transitional ecosystems. *Transitional Waters Bulletin*. 2013, Vol. 7, No. 1, p. 43-63. ISSN 1825-229X.
- [27] Rumbaugh, J., Jacobson, I., Booch, G. Unified Modeling Language Reference Manual. The. Pearson Higher Education. 2004. ISBN:0321245628.
- [28] San Gil, I., Baker, K., Campbell, J., Denny, E. G., Vanderbilt, K., Riordan, B., Koskela, R., Downing, J., Grabner, S., Melendez, E., Walsh, J. M., Kortz, M., Conners, J., Yarmey, L., Kaplan, N., Boose, E. R., Powell, L., Gries, C., Schroeder, R., Ackerman, T., Ramsey, K., Benson, B., Chipman, J., Laundre, J., Garritt, H., Henshaw, D., Collins, B., Gardner, Ch., Bohm, S., O'Brien, M., Gao, J., Sheldon, S., Lyon, W., Bahauddin, D., Servilla, M., Costa, D., Brunt, J. The Long-Term Ecological Research community metadata standardisation project: a progress report. *International Journal of Metadata, Semantics and Ontologies*. 2009, Vol. 4., No. 3, p. 141-153. ISSN 1744-2621, e-ISSN 1744-263X.
- [29] San Gil, I., Hutchinson, V., Frame, M., Palanisamy, G. Metadata Activities in Biology. *Journal Library Metadata*. 2010, Vol. 10, No. 2-3, p. 99-118. ISSN 1938-6389, e-ISSN 1937-5034.
- [30] The Knowledge Network For Biocomplexity. Ecological Metadata Language (EML) Specification. 2014.