A Holistic, Semantics-aware Approach to Spatial Data Infrastructures

Cristiano Fugazza, Monica Pepe, Alessandro Oggioni, Fabio Pavesi and Paola Carrara
IREA-CNR, via Bassini 15, 20133 Milano, Italy

Keywords: Spatial Data Infrastructures, Semantic Web, Marine Research, RDF.

Abstract: We present a novel approach to the management of Spatial Data Infrastrutures that leverages semantics-aware context information to model the distinct aspects involved in the management of geospatial data. RDF-based schemata are employed for encoding information about the user community, the terminologies in use in a specific research domain, gazetteer information representing the physical landscape underpinning data and, last but not least, resource metadata. The data structures are then interconnected to enable seamless exploitation for metadata creation and resource discovery, which we demonstrate through a worked-out example of SPARQL query on RDF graph data. The methodology is being applied by the National Research Council of Italy (CNR) to support creation of a distributed infrastructure for marine data in the context of the RITMARE Flagship Project.

1 INTRODUCTION

Spatial Data Infrastructures (SDIs) represent one of the most challenging areas in data management. In fact, the variety of data and metadata formats, the different workflows in use, and the manifold expectations from the user community make it difficult to apply widely acknowledged design patterns. Moreover, the ever-changing landscape of data sources - think of the emergence of sensor data (Na and Priest, 2007) - and the evolution of data access methodologies - e.g. the Linked Open Data (LOD) movement (Dodds and Davis, 2012) - prevent data management practices to easily coalesce into generally applicable standards.

Recourse to semantics in the management of geospatial data is typically intended to address heterogeneity in resource description, to achieve multilingualism in resource discovery functionalities, and in general to compensate for the lack of the efficient search paradigms of generalized information retrieval. However, no existing infrastructure employs semantics as the *trait d'union* between the distinct components of the context information involved in SDIs. Also, no state-of-the-art solution applies this approach throughout the life cycle of geospatial information, from metadata production to resource retrieval. Instead, it is our opinion that only by adopting an all-round approach encompassing all the aspects of SDIs it is possible to fully exploit the potential of the expressive descriptions enabled by ontologies, thesauri, and RDF-based data structures in general[1].

With this in mind, IREA-CNR[2] tackled the challenge of providing an infrastructure to the RITMARE Flagship Project[3], which aims at bringing all the distinct contributions to Italian marine research under the same umbrella. In doing this, RDF-based data structures are exploited to model the distinct aspects of the SDI. These range from the categorization of researchers and institutions involved in the project to the collection of the terminologies in use by the distinct communities; from the organization of a gazetteer system supporting geospatial information retrieval to the encoding of metadata as Linked Data records. The latter component will contribute to the LOD cloud[4] with a "bubble" constituted by the metadata collected by RITMARE.

In this paper, we report on the activities that are carried out for the establishment of the RITMARE infrastructure with regard to semantics. Section 2 describes the project's generalities and presents some related work that is relevant to our research. Section 3 describes the RDF-based data structures that support the front- and back-office functionalities of the infras-

[1] Resource Description Framework: http://www.w3.org/RDF/

[2] Institute for Electromagnetic Sensing of the Environment - National Research Council of Italy: http://irea.cnr.it/

[3] RITMARE (Ricerca ITaliana per il MARE - Italian research for the sea): http://www.ritmare.it/

[4] Linked Open Data cloud: http://lod-cloud.net/

Fugazza C., Pepe M., Oggioni A., Pavesi F. and Carrara P..
A Holistic, Semantics-aware Approach to Spatial Data Infrastructures.
DOI: 10.5220/0004997603490356
In *Proceedings of 3rd International Conference on Data Management Technologies and Applications* (DATA-2014), pages 349-356
ISBN: 978-989-758-035-2

tructure. Section 4 provides a preliminary evaluation of the approach that is presented. Finally, Section 5 draws conclusions and outlines future work.

2 CONTEXT

The building of the data sharing infrastructure in RITMARE (a Flagship Project by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca) requires integration of all state-of-the-art contributions to Italian marine research into a coherent SDI. A coarse-grained categorization of SDIs distinguishes between centralized and decentralized infrastructures, according to whether data and metadata is stored in a single repository or distributed among the distinct data providers. The RITMARE infrastructure belongs to the second kind, featuring:

- a set of peripheral nodes that expose standards-compliant metadata and services;
- a centralized catalog service that provides a single point of access to the resources made available by the project as a whole.

The project involves a heterogeneous set of data providers (public research bodies and inter-university consortia) as well as a variety of stakeholders (public administrations, private companies, and citizens). These entities envisage a varied corpus of heterogeneous data, metadata, workflows, and requirements. Besides this, data providers feature different degrees of maturity with regard to the provisioning of resources according to the mandated standards. In fact, the RITMARE SDI is bound to the requirements set by the INSPIRE (INfrastructure for SPatial InfoRmation in Europe) Directive (European Commission, 2007), as well as by RNDT[5], the Italian enforcement of the former.

The development activity can be roughly divided into three incremental phases. The first one consists of empowering the data providers with standards-compliant services for the provisioning of geospatial data. This has been achieved by providing a virtual appliance, a “starter kit” (Fugazza et al., 2014), that is capable of kickstarting an autonomous node in the SDI for the collection, annotation, and deployment of both geographic and sensor data. The data sources that already expose standard services will be integrated in the second phase. Finally, the third phase will consist of the building of the centralized portal integrating the distinct peripheral nodes and exposing the metadata records as Linked Data.

In the first two phases, the data structures described in this work allow to weave into the workflow for resource registration advanced functionalities for the enrichment of metadata. In the third phase, the centralized repository will exploit them for articulating resource discovery in a more effective way as regards to existing *geoportals* (the front-end of SDIs).

2.1 Related Framework

INSPIRE-compliant, XML-based metadata and OGC-based services (European Commission, 2008; IOC Task Force for Network Services, 2011a; Initial Operating Capability Task Force, 2012; IOC Task Force for Network Services, 2011b) seem to find it hard to keep pace with more recent LOD initiatives, even those sponsored by the EU itself, such as its Open Data Portal[6]. However, an extension to the *DCAT*[7] *application profile for data portals in Europe* (European Commission, 2013) is expected to ease transition to RDF-based metadata representation for INSPIRE metadata. The paper (Fugazza and Luraschi, 2012) proposes an indexing process for INSPIRE metadata based on SKOS thesauri. The methodology is applied to harvested metadata records and hence is likely to be error-prone, while our approach in this worst-case scenario (that is, when harvesting metadata from existing infrastructures) envisages supervised amelioration of metadata in order to obtain RNDT-compliant metadata, providing a more reliable indexing as by-product.

In extra-European contexts, such as the GEO System of Systems[8], semantics is intended as the primary means to overcome the generalized hetereogeneity of formats (Khalsa et al., 2007; Nativi et al., 2011). In particular, in (Santoro et al., 2012) the authors propose a query expansion mechanism that is applied to the search string entered by the user, rather than to the metadata records themselves. On the one hand, this approach is valuable because it is applicable when metadata records are not known in advance. On the other hand, precision and recall in discovery is inevitably hampered by the scarce information associated with the user executing the search (e.g., the language the query is expressed in is not known in advance).

Another important issue that we wanted to address is the inclusion in our marine infrastructure of Sen-

[5]Repertorio Nationale dei Dati Territoriali: http://www.rndt.gov.it

[6]EU Open Data Portal: https://open-data.europa.eu

[7]Data Catalog Vocabulary (DCAT): http://www.w3.org/TR/vocab-dcat/

[8]GEOSS: https://www.earthobservations.org/geoss.shtml

sor Web Enablement[9] (SWE) functionalities, particularly the Sensor Observation Service (SOS) component (Oggioni et al., 2014; Jiang et al., 2013). Ontology-based representation of sensors and observations is featuring noteworthy results, as in (Barnaghia et al., 2011; Compton et al., 2012; Wang, 2012; Taylor et al., 2011; Taylor and Leidinger, 2011), that are inspiring our approach to the representation of this category of data sources within our RDF-based SDI. However, the challenge of harmonizing sensor networks, particularly observations (that is, the actual data produced by sensors), with traditional geographic data remains. Ideally, one should transparently access sensor data alongside the other categories of geospatial data and exploit the specificities, characteristics, and history of sensors (e.g. measured parameters, gain, accurancy, offset, calibration features) when needed. Instead, Sensor Web development relies on specific standards and practices that make it difficult to achieve this degree of integration (Havlik et al., 2011).

3 THE RITMARE LOD BUBBLE

Figure 1 depicts the individual components of the RDF data structures that are being populated for developing the RITMARE infrastructure. The size of each circle is proportional to the expected number of entities in each data structure. Also, in order to exploit these data structures in a seamless way, it is necessary to provide the appropriate relations between entities from distinct schemata. The following of this Section will describe the specificities of the individual components with regard to i) the schemata that are employed, ii) the data that are stored, and iii) the relations that are created for connecting the distinct domains. Section 3.5 provides some implementation details on the two data bases that are conjunctively employed. Finally, Section 3.6 describes an actual discovery scenario, which is a worked-out example of geospatial query interrogating all the data structures to obtain a list of prospective results.

3.1 Project Description

Researchers and institutions involved in the project are modeled as Friend Of A Friend (FOAF) data structures[10]. The rationale for this approach is the following: Firstly, FOAF is a widely-acknowledged data format for establishing a Web of Trust among the participants to the project. Also, FOAF descriptions can refer to terms in controlled vocabularies representing the research field of individual persons and whole institutions: As an example, Figure 1 shows the FOAF entity corresponding to researcher "Monica Pepe" linked to the SKOS concept corresponding to "Geophysics" by property *topic_interest* from the FOAF vocabulary. This semantic link is the primary means we adopted to provide advanced metadata creation functionalities, enable profiled discovery of resources, and tailor the selection of components to be included in the infrastructure's geoportal on the user's specificities.

The categorization of the entities involved in RITMARE has been derived from documents provided by the administration of the project as a spreadsheet. The documents have been exported as XML and processed with an XSLT stylesheet in order to obtain the raw RDF/XML data structures. In this phase, human supervision was required in order to prune duplicate individuals, remove typos, etc. Finally, the workflow for creating links between FOAF entities and terms of the knowledge base is twofold: For testing purposes, an ad-hoc backoffice application has been created; instead, when the centralized geoportal will be available, the research field associated with a given individual will be one of the preferences that the user can specify and modify.

3.2 Knowledge Base

The second component is constituted by a collection of terminologies in the SKOS format. These include the multilingual codelists that are employed in the creation of INSPIRE- and RNDT-compliant metadata, a selection of thesauri whose terms can be used as keywords, and the categorization of the RITMARE research areas. An activity that may be required in the future is the harmonization of distinct controlled vocabularies, that is, the drawing of correspondences between related terms in independent terminologies. For instance, in Figure 1 the SKOS concepts corresponding to "Elevation" and "Altitude" are related to each other by semantic property *closeMatch* from the SKOS ontology. This aspect is a fundamental enabling factor for query expansion functionalities that, given the inter-disciplinary vocation of RITMARE, may be an important component for the usability of the infrastructure.

Whenever a suitable SKOS representation is available, thesauri have been downloaded from the authoritative sources and inserted in the knowledge base with little or no modification. Instead, most of the

[9]Sensor Web Enablement: http://www.opengeospatial.org/ogc/markets-technologies/swe

[10]FOAF Vocabulary Specification 0.99: http://xmlns.com/foaf/spec/

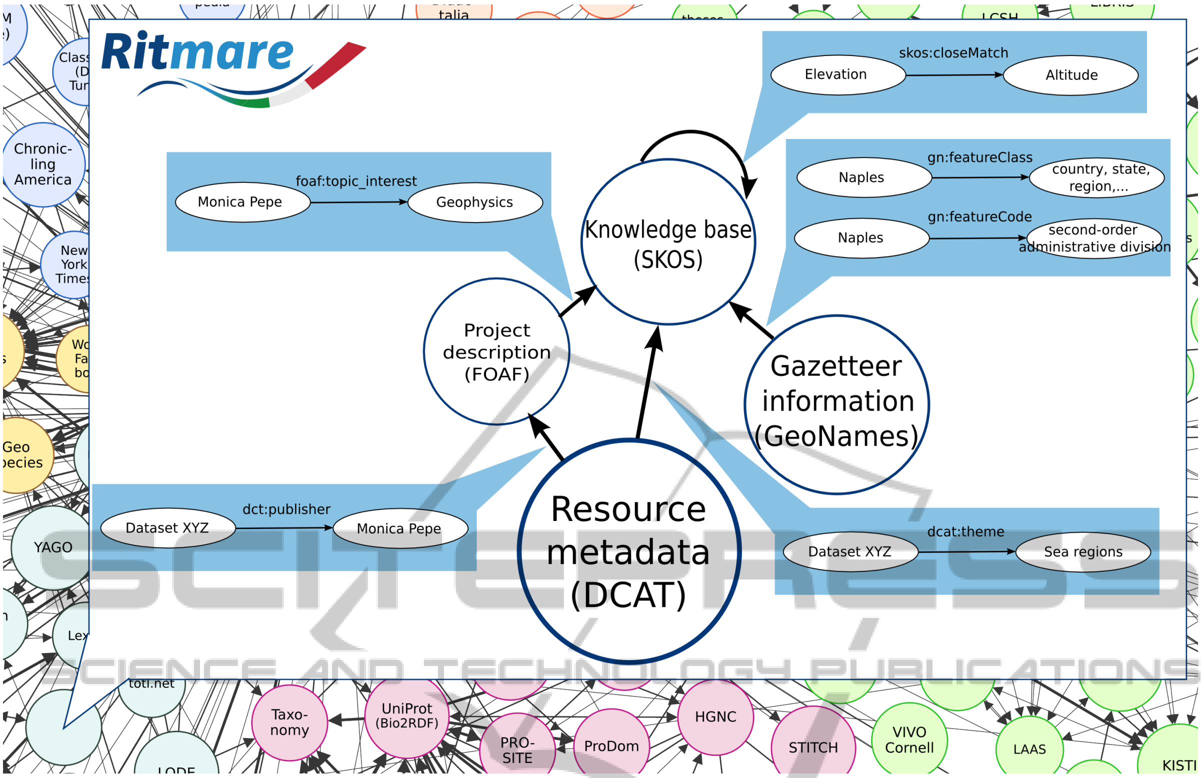

Figure 1: Components of the RITMARE LOD bubble.

codelists that are used in INSPIRE metadata had to be created from scratch by referring to INSPIRE and ISO documentation. The same for the list of research fields in RITMARE. As for the selection of thesauri to be supported for keyword selection, consultation with domain experts inside and outside of the project recommended some of the thesauri provided by SeaDataNet[11] (particularly, those categorizing observation parameters, units of measure, sensor platforms, and sensor instruments) and the GCMD Science Keywords[12] provided by NASA. We deliberately avoided including some thesauri that are typically used in the geospatial domain, such as GEMET[13], as they were deemed as too general by this specific user community. However, since referring to thesauri is a fundamental component in RITMARE metadata, we encourage the project partners to provide feedback on new terminologies to be included in the knowledge base.

[11]SeaDataNet BODC webservices: http://seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx

[12]Global Change Master Directory: http://gcmd.nasa.gov/

[13]GEneral Multilingual Environmental Thesaurus: http://www.eionet.europa.eu/gemet/

3.3 Gazetteer Information

Another adopted data structure is a fine-grained representation of the geographic features that may be referred to in user queries. For the time being, we only included some of the toponyms provided by GeoNames[14] in order to be able to express query patterns like "100km West of Naples" and reduce recourse to maps, bounding boxes, and other widgets that typically distinguish geographic data discovery from generalized search. Gazetteer data structures are already referring to controlled vocabularies, also included in the knowledge base, for categorizing toponyms. As an example, Figure 1 shows two SKOS concepts (described by literals "country, state, region, ..." and "second-order administrative division") that are referred to by the toponym corresponding to "Naples" as values for, respectively, properties *featureClass* and *featureCode* defined by the GeoNames ontology. In the future, it is possible that the GeoNames ontology could be complemented by (or directly repurposed for expressing) a custom RITMARE maritime gazetteer.

The GeoNames RDF dump (around 3GB) is too large for direct usage and too large for DOM and

[14]GeoNames: http://www.geonames.org/

XSLT processing tools. We then developed a Python script that extracted from the original file the Italian toponyms. The thesaurus categorizing GeoNames' toponyms also allowed to prune the toponyms that were clearly of no interest to the project (e.g., those corresponding to hotels, parks, etc.) so that the data actually inserted in the triple store was just around 50MB.

3.4 Resource Metadata

Finally, the largest proportion of the RDF data employed by the RITMARE infrastructure is constituted by resource metadata harvested from the peripheral nodes. For their encoding, we rely on the Data Catalog Vocabulary (DCAT) and its profile tailored on the requirements of the EU Open Data Portal. Until a proper extension for expressing INSPIRE metadata is made available, we rely on the RDF representation of ISO 19115 metadata provided by the Australian CSIRO[15]. In the example in Figure 1, the dataset "Dataset XYZ" has been related to the INSPIRE Theme "Sea regions" and to the publisher "Monica Pepe" by properties *theme* and *publisher* from the DCAT vocabulary, respectively.

The enrichment of metadata records may follow two distinct directions; in both cases, we essentially annotate metadata items with the identifiers (the URIs) of the RDF entities described above. When these entities are already provided by sources in the LOD cloud (e.g., GeoNames toponyms), they refer to the corresponding Linked Data structures. The metadata produced via the tool that we developed in the first phase of the project (see Section 2) already contain this information. Otherwise, when metadata is harvested from an existing data source, indexing techniques will be applied to complement the language-dependent textual descriptions with the corresponding language-neutral identifiers.

3.5 Implementation Details

The first three components that are described above are stored in an instance of the Virtuoso Universal Server[16] (although only the triple store capabilities are employed). This platform provides all the functionalities that are required for the management of RDF data and is accessed by the distinct components of the architecture through the SPARQL[17] endpoint that is provided. The choice of the solution to be used for the deployment of the fourth component, the metadata of the resources aggregated by the infrastructure, is currently not definitive. For the time being, we are evaluating CKAN[18], the mainstream solution for the implementation of open-data portals. Besides support for the provision of resource metadata as Linked Data, this platform natively exploits the DCAT metadata format, which is the basis for the RDF-based resource description that we employ.

The motivations for using two distinct platforms are manifold. Firstly, Virtuoso offers a wide set of tools specifically aimed at database administration and this feature has been dramatically important in the beginning of the project, when multiple schemata, thesauri, and prospective data representations were evaluated. On the other hand, CKAN may constitute a more scalable solution for the category of data that is likely to grow in size, that is, resource metadata. Secondly, the distinct components may feature a different status with regard to access control or authoritativeness of the source. For instance, user details involve privacy issues that we wanted to keep under control. Also, some components that constitute the knowledge base (say, the SeaDataNet thesauri) already have their authoritative sources and then providing them through a second endpoint would be redundant. Finally, in the future, some components of the context information described in this Section, such as thesauri and code lists, could be drawn directly from the respective SPARQL endpoints or Linked Data facilities. For the time being, these data structures are materialized in the local repository for performance and direct control.

3.6 Worked-out Example

The combined expressiveness of the four components described above support the discovery capabilities exemplified in this Section, in which the metadata catalog is interrogated by means of the SPARQL query language. Queries are composed on the basis of parameters specified by the user through a discovery interface (a web form) hinged on four elements:

WHO: This parameter is used for ranking results according to the properties *topic_interest* that are specified for a given researcher (if authenticated). The effect of this parameter is otherwise transparent to the end user.

WHAT: Describes the observation parameter of interest to the user. When no match is found in the

[15]ISO 19115 RDF representation: http://def.seegrid.csiro.au/static/isotc211/iso19115/2003/

[16]Virtuoso: http://virtuoso.openlinksw.com/

[17]SPARQL Protocol And RDF Query Language: http://www.w3.org/TR/rdf-sparql-protocol/

[18]CKAN The open source data portal software: http://ckan.org/

Listing 1: Example of SPARQL query for discovery of datasets.

```
1  PREFIX foaf:  <http://xmlns.com/foaf/0.1/>
2  PREFIX vcrd:  <http://www.w3.org/2006/vcard/ns#>
3  PREFIX skos:  <http://www.w3.org/2004/02/skos/core#>
4  PREFIX dcat:  <http://www.w3.org/ns/dcat#>
5  PREFIX dtc:   <http://purl.org/dc/terms/>
6  PREFIX ext:   <http://def.seegrid.csiro.au/isotc211/iso19115/2003/extent#>
7
8  SELECT ?dataset ?title ?abstract
9  WHERE {
10   ?dataset rdf:type dcat:Dataset ;
11            dct:title ?title ;
12            dct:description ?abstract;
13            dcat:theme <http://vocab.nerc.ac.uk/collection/P02/current/PSAL/> ;
14            dcat:publisher ?publisher .
15   ?publisher vcrd:org <http://ritmare.it/rdfdata/project#ISMAR> .
16   {
17     ?dataset ext:spatialExtent ?bbox .
18     ?bbox  ext:northBoundLatitude ?nlat ;
19            ext:southBoundLatitude ?slat ;
20            ext:westBoundLongitude ?wlong ;
21            ext:eastBoundLongitude ?elong .
22     FILTER( (?nlat > 40.9642) && (?slat < 40.8528) &&
23             (?wlong > 13.1573) && (?elong < 12.9567) )
24   }
25 }
```

related controlled vocabulary, a free-text search is triggered by the application.

WHERE: Specifies the geographic context of the search as a free-text parameter to be interpreted by the application.

WHEN: Determines the temporal extent of interest to the user.

Also, the user is given the option to refine the search parameters by specifying facet values for some metadata items. These are the fields whose possible values are known in advance (such as the INSPIRE Theme, the topic category, the language used in metadata, etc.) or that can be statically computed and clustered. In the example, the user has selected, in a text field that is autocompleted by leveraging the SeaDataNet thesaurus for observation parameters, the term corresponding to “Salinity of the water column”. She also specified the free-text value “100km west of Naples” as the geographic location of interest. It is up to the application to guess (in a Google Maps-style) the distance, direction, and origin expressed in the text field. Moreover, she also selected, among the data providers in project RITMARE, the facet value corresponding to the “ISMAR” institute.

Executing SPARQL queries amounts to searching the graph data whose components have been introduced in Section 3 for *triple patterns*. These are sequences of subject-property-object patterns (subjects and objects like the ellipses in Figure 1, the properties like the directed arcs) inducing a graph that can then be matched against the data graph. User queries are translated into SPARQL queries: In particular, the parameters described above translate into Listing 1.

The explanation of the listing is the following. Lines 1-6 define namespace prefixes that allow to shorten the triple patterns in the listing. Namely, the prefixes identify the namespaces of the following vocabularies: FOAF, vCard (used in FOAF), SKOS, DCAT, and Dublin Core Terms (used in DCAT). Also, prefix *ext* identifies the data structures that compensate for the lack of proper DCAT data structures for expressing some INSPIRE metadata items. The SELECT clause on line 8 defines the three variables that shall be returned by the query, that is, the URI, title, and abstract of the resources matching the query. The WHERE clause defines the triple patterns that are searched in the RDF data structures of RITMARE and which data shall be bound to variables. Lines 10-12 retrieve the data requested in the projection clause on line 8, while lines 13-15 is where the actual matching starts. Specifically, the first line matches metadata records that are related to the observation parameter that has been requested by the user. The remaining two require that the point of contact that is provided for data belongs to the institute that has been selected by the user for the corresponding facet.

Finally, lines 17-23 express the tricky part of the query, that is, the matching of the bounding box specified by the user with the free-text parameter “100km west of Naples”. Here, the geographic coordinates corresponding to Naples (40.88333N, 14.41667E according to the toponym specified by GeoNames) are applied the offset corresponding to 100km westwise (obtaining 40.90802N, 13.05750E). Then, a bounding box has been computed from this position (the four values in the FILTER clause on lines 22 and 23 of Listing 1). Here, we assume that the user is fine with the default dimensions of the bounding box provided by the discovery application, 16km wide and 11km high. The (possibly multiple) geographic extents of each candidate dataset are matched against the search bounding box. Here, we assume the more conservative approach, requiring the latter to be fully contained in the former. However, a range of different interpretations are possible. Particularly, in the future we will aim at providing the user with the capability of modulating the degreee of overlapping of bounding boxes according to the number of results that would be returned by the query.

4 DISCUSSION

In the current phase of the project, when the centralized geoportal is still work in progress, an overall evaluation of the proposed solution is not easy. However, we are already collecting feedback on the capacity building of data providers envisaged in the first phase of the project (see Section 2), that is, we can report on the efficacy of our “starter kit”. In particular, we can evaluate the features that we are providing for the assisted editing of metadata, which are heavily dependent on the data structures described in this paper.

With respect to these, recourse to the RDF data structures presented in Section 3 allows to dramatically reduce the number of the data items to be actually provided by the user. As an example, the RNDT specification is featuring 34 mandatory metadata items, too high a number for an accurate editing of metadata records. Moreover, some of these, such as the 4 distinct contact points to be provided, are composite items that require more than one value to be entered by the user. Instead, by combining i) the automatic assignment of metadata items, ii) the provision of default values extracted from the dataset itself, and iii) the exploitation of the RDF information described in this paper, the number of required values can be narrowed down to 10.

Particularly, by referring to the project structure, the editing interface is capable of providing data items that are otherwise tedious to provide or error-prone. Also, autocompleting keywords from the selected thesauri configures a win-win situation where the user is relieved of the burden of writing keywords in their entirety and the application is capable of associating unique identifiers with the keywords that are provided. These advanced editing capabilities can be applied to generic metadata schemata encoded as XML. In fact, we currently support the editing of sensor metadata in the SensorML format with the same autocompletion facilities of RNDT metadata. Also, support for sensor metadata in the alternative *FL (Starfish Fungus Language) format (Simonis and Malewski, 2011) is under testing.

5 CONCLUSIONS

This work presents our approach to semantic characterization of the data structures that are involved in the management of SDIs. We believe that, by creating the appropriate relations between data items encoded according to the heterogeneous RDF schemata presented in Section 3, it is possible to provide end users with advanced functionalities. These span from back- to front-office capabilities, from metadata creation by the peripheral nodes of the RITMARE infrastructure to resource discovery by the centralized catalog, and permeate the whole life cycle of geospatial resources as semantic annotations to INSPIRE-compliant metadata.

The proposed methodology is supporting the development of the user-oriented infrastructure of the RITMARE Flagship Project and constitutes the first example of decentralized management of Italian marine research data. The novelty of the approach also resides in the all-round application of semantics in the modeling of the data structures that describe resources, end users, research fields, and physical locations. The example provided in Section 3.6 only scratches the surface of the possible applications of this multi-tiered knowledge base in the enactment of Spatial Data Infrastructures.

ACKNOWLEDGEMENTS

The activities described in this paper have been funded by the Italian Flagship Project RITMARE.

REFERENCES

Barnaghia, P., Ganza, F., and Abangara, H. (2011). Sense2Web: A Linked Data Platform for Semantic

Sensor Networks. *semantic-web-journal.net*, 0(0):1–11.

Compton, M., Barnaghi, P., Bermudez, L., Garca-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Phuoc, D. L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., and Taylor, K. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17(0):25 – 32.

Dodds, L. and Davis, I. (2012). Linked Data Patterns, http://patterns.dataincubator.org/book/.

European Commission (2007). "Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)" Directive 2007/2/EC.

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.

European Commission (2013). DCAT Application Profile for data portals in europe, https://joinup.ec.europa.eu/asset/dcat_application_profile/home.

Fugazza, C. and Luraschi, G. (2012). Semantics-Aware Indexing of Geospatial Resources Based on Multilingual Thesauri : Methodology and Preliminary Results. *International Journal of Spatial Data Infrastructures Research*, 7:16–37.

Fugazza, C., Menegon, S., Pepe, M., Oggioni, A., and Carrara, P. (2014). The RITMARE Starter Kit: Bottom-up capacity building for geospatial data providers. In *9th International Conference on Software Paradigm Trends (ICSOFT-PT)*.

Havlik, D., Schade, S., Sabeur, Z. a., Mazzetti, P., Watson, K., Berre, A. J., and Mon, J. L. (2011). From Sensor to Observation Web with environmental enablers in the Future Internet. *Sensors (Basel, Switzerland)*, 11(4):3874–3907.

Initial Operating Capability Task Force (2012). Technical guidance for the implementation of inspire download services.

IOC Task Force for Network Services (2011a). Technical guidance for the implementation of inspire discovery services.

IOC Task Force for Network Services (2011b). Technical guidance for the implementation of inspire view services.

Jiang, Y., Guo, Z., Hu, K., Shen, F., and Hong, F. (2013). Using Sensor Web to Sharing Data of Ocean Observing Systems. In Wang, R. and Xiao, F., editors, *Advances in Wireless Sensor Networks*, pages 137–156. Springer-Verlag Berlin Heidelberg, Berlin Heideberg.

Khalsa, S. J. S., Nativi, S., Ahern, T., Shibasaki, R., and Thomas, D. (2007). The GEOSS interoperability process pilot project. In *IGARSS*, pages 293–296.

Na, A. and Priest, M. (2007). OGC Sensor Observation Service (OGC 06-009r6). OpenGIS Implementation Standard.

Nativi, S., Craglia, M., Vaccari, L., Santoro, M., and Fugazza, C. (2011). Searching the New Grail: Inter-Disciplinary Interoperability. In Claramunt, C., Levashkin, S., and Bertolotto, M., editors, *GeoSpatial Semantics, Proceedings of the 4th International Conference, Geos 2011*, number 1, pages 1–15. Springer-Verlag.

Oggioni, A., Bastianini, M., Carrara, P., Minuzzo, T., and Pavesi, F. (2014). Sensing Real-time Observatories in Marine Sites A Proof-of-Concept. In *SensorNets 2014, 3rd International Conference on Sensor Networks*, pages 111–118, Lisbon Portugal 7-9 January 2014.

Santoro, M., Mazzetti, P., Nativi, S., Fugazza, C., Granell, C., and Díaz, L. (2012). Methodologies for augmented discovery of geospatial resources. In Díaz, L., Granell, C., and Huerta, J., editors, *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications*, chapter 9.

Simonis, I. and Malewski, C. (2011). *FL Starfish Fungus Language for sensor description. OGC discussion paper 11-058r1.

Taylor, K., Compton, M., and Laurent, L. (2011). Semantically-enabling the web of things: The W3C semantic sensor network ontology. In *Proc. of 5th eResearch Australasia Conference*.

Taylor, K. and Leidinger, L. (2011). Ontology-driven complex event processing in heterogeneous sensor networks. In Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., Leenheer, P., and Pan, J., editors, *The Semanic Web: Research and Applications*, volume 6644 of *Lecture Notes in Computer Science*, pages 285–299. Springer Berlin Heidelberg.

Wang, P. (2012). *Semantically enabling next generation environmental informatics portals*. PhD thesis, Rensselaer Polytechnic Institute.