

Article

Web-Scale Normalization of Geospatial Metadata Based on Semantics-Aware Data Sources

Cristiano Fugazza ^{1,*}, Paolo Tagliolato ^{1,2}, Luca Frigerio ¹ and Paola Carrara ¹

¹ Institute for Electromagnetic Sensing of the Environment, National Research Council (IREA-CNR), 20133 Milan, Italy; tagliolato.p@irea.cnr.it (P.T.); frigerio.l@irea.cnr.it (L.F.); carrara.p@irea.cnr.it (P.C.)

² Institute of Marine Science, National Research Council (ISMAR-CNR), 30122 Venice, Italy

* Correspondence: fugazza.c@irea.cnr.it; Tel.: +39-02-2369-9456

Received: 29 September 2017; Accepted: 2 November 2017; Published: 13 November 2017

Abstract: Geospatial metadata are largely denormalized inasmuch as resource descriptions typically accommodate property values as plain text. Hence, it is not possible to bring multiple references to the same entity (say, a keyword from a controlled vocabulary) under the same umbrella. This practice is ultimately the main source for the heterogeneities in metadata descriptions by which geospatial discovery is hampered. In this paper, we elaborate on ex-post semantic augmentation of metadata, a technique generally referred to as semantic lift, which complements our previous research on semantic characterization of metadata via transparent association of uniform resource identifiers with metadata items at editing time. The latter is accomplished by means of a template-based metadata editor that can be tailored to any XML-based metadata schema. By repurposing the template language previously defined for metadata editing, we broaden the expressiveness of the former and integrate heterogeneous, XML-based resource descriptions in our semantics-aware metadata management workflow. URI-based indirection in metadata provision not only entails normalization of individual information items and allows one to overcome the aforementioned heterogeneities, but also elicits decentralized, multi-tenanted management of metadata.

Keywords: spatial data infrastructures; geospatial metadata; RDF; semantic lift

1. Introduction

The transition from GIS to SDI has been fostered by advancements in IT—primarily, the advent of the Internet as a medium for communication—and policy requirements demanding for a better national and transnational management of geospatial resources [1]. Two closely-related consequences of this transition were: (i) the enlargement of the intended audience of a GI product, now spanning a worldwide network of potential users; and (ii) the heterogeneity of application domains and the degrees of expertise of the latter. Interoperability and inter-disciplinary aptness of geospatial resources has thus become an inalienable requirement [2,3].

A number of activities tackled the challenge of geospatial data interoperability in the last decade, both at the European and the global level, such as INSPIRE [4–6] and GEOSS [7,8]. These initiatives primarily addressed this issue from a technical viewpoint, although GEOSS inflected the subject so as to identify the need for geospatial data interoperability in specific societal benefit areas. Moreover, the notion of digital Earth [9,10] emphasized the breadth of this challenge and the variety of stakeholders that can take advantage of harmonized management of geospatial data.

Interoperability emerges as an even more pivotal issue in research contexts, where tracking the lineage of resources and supporting their reuse is essential to the identification and exploitation of reference data sources. Figure 1 portrays the data management workflow as conceived by the ENVRI Community (<http://envri.eu/>) [11]. Albeit, some CoP may raise concerns about the interpretation of some of the phases in the picture, it represents a possible workflow articulating the diverse activities

that contribute to effective data management. In particular, we stress the important role of metadata in the curation and publishing phases.

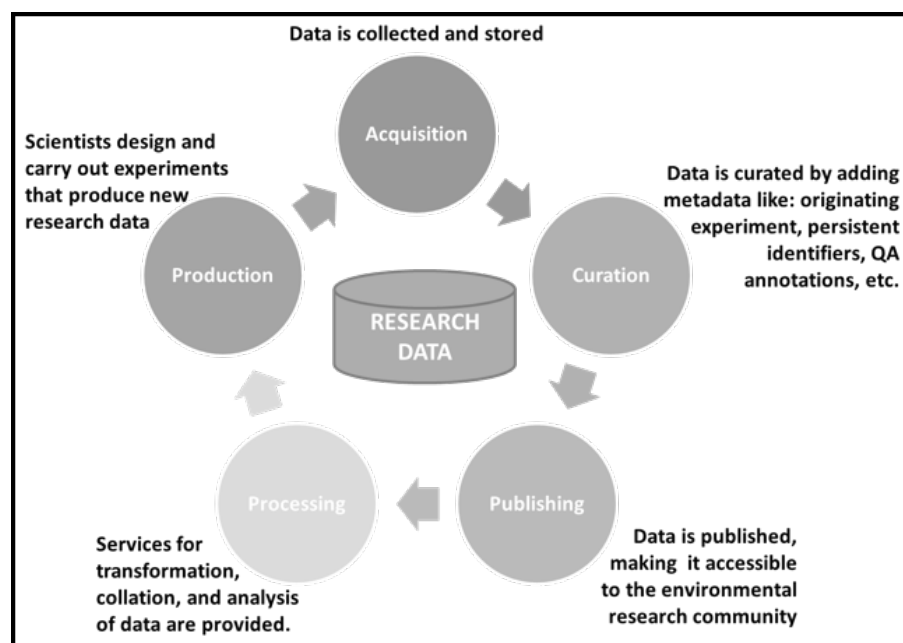


Figure 1. Data management workflow as conceived by the ENVRI Community.

Whereas all phases in this ideal walkthrough can harness the broad range of services standardized by OGC (Open Geospatial Consortium, <http://www.opengeospatial.org/>), annotation and discovery of datasets are intrinsically hampered by semantic heterogeneities that are even more apparent in multilingual contexts such as that of the EU. Focusing on the issues entailed by the discovery of geospatial resources, our vision is that semantic enrichment of metadata is pivotal in solving most heterogeneity problems. Section 2 is going to substantiate this assertion by providing three use cases.

In [12], Bishr defines six levels of interoperability for SDIs and a more coarse-grained categorization that distinguishes between syntactic, schematic and semantic heterogeneities. Oukssel and Sheth [13] expand the first two elements in this categorization while Ramakrishnan et al. [14] elaborate on the last one. Referring to the categorization in [12], the aforementioned OGC standards help address two categories of heterogeneities out of three, such as in the implementation presented by Nativi and Mazzetti in [15]. Instead, our work addresses semantic heterogeneity (specifically, referring to [14], it addresses the conferment of implicit semantics to metadata).

Our metadata management is hinged on semantic characterization of metadata via transparent association of URIs [16] with metadata items; this is accomplished in a two-fold fashion. Ex-ante semantic augmentation is enabled by EDI [17], a template-based metadata editor that can be tailored to any XML-based metadata schema. Instead, in this paper, we elaborate on ex-post semantic augmentation of metadata, a technique generally referred to as semantic lift, via a FOSS product named Liftboy and made available in our GitHub repository (<https://github.com/SP7-Ritmare>). Specifically, we repurpose and extend the template language defined for EDI, and among the many metadata schemas adopted in geospatial resource management, we address the ISO TC 211 standard series for geographic information [18–20] for testing purposes. URI-based indirection in metadata provision not only allows one to overcome the aforementioned heterogeneities, but also elicits a novel paradigm in metadata management.

In fact, the technical advantages described so far subtend two important theoretical issues in (not only geospatial) metadata management. The first relates to normalization of metadata records, that is ensuring that multiple references to the same information item map to the same digital construct

(in our framework, via URIs) instead of being duplicated. By doing this, it is possible to improve the consistency of resource descriptions and automatically update individual property values [21]. Closely related to normalization, the second issue involves metadata delegation, that is allowing that some property values in a record may be provided (and maintained) by a third party. Letting aside the technical requirements that are necessary to implement such a delegation principle (requirements that are met by our implementations), one should also be aware of the necessary paradigm shift from the notion of data structures that are single-handedly maintained by the institution providing a given resource to that of distributed metadata.

It is not apparent how current, centralized metadata management practices could easily migrate to this new paradigm. In fact, albeit that semantic characterization of metadata items is realized by our applications, web-scale management of this information requires design, communication and governance. The Open Annotation Community Group (<https://www.w3.org/community/openannotation/>) by W3C (formerly Web Annotation Group) provides an ecosystem of data formats and protocols for managing RDF-based, decentralized annotation of digital resources. Its scope is far wider than ours, but this initiative can nevertheless provide the missing components in our metadata management framework. In particular, in this work, we show that the semantic information distilled from XML-based metadata can be expressed as WADM annotations [22], one of the products of the aforementioned working group.

This paper is organized as follows. Before describing how semantic lift can be achieved by means of the Liftboy application, in Section 2, we motivate why semantic lift is necessary at all and outline the state of the art. Then, in Section 3, we describe the essentials of the template language we created, present Liftboy, the application we developed for achieving semantic lift of resource descriptions, and detail the workflow that is followed for producing semantics-aware metadata out of traditional, XML-based ones. In this section, we also discuss the profound differences between ex-ante and ex-post semantic augmentation of metadata with a use case addressing RNDT [23], the Italian transposition of INSPIRE (i.e., ISO 19115/19119) metadata. Section 4 describes serialization of the semantic information produced by Liftboy as WADM annotations. Section 5 revisits the use cases presented in Section 2 and presents worked-out examples of discovery based on the lifted metadata. Finally, Section 6 draws conclusions and outlines the integration of Liftboy in our metadata management framework.

2. Semantic Characterization of Geospatial Metadata

As already mentioned, discovery of geospatial resources has always constituted a daunting task because of the pivotal role of metadata. In fact, the aforementioned heterogeneity issues, together with the scarcity and inaccuracy of metadata, hamper the effective search for the intended resources over the web. Data schema mapping via XSLT [24], also supported by INSPIRE Data Specifications (<https://inspire.ec.europa.eu/data-specifications/2892>), already features virtuous examples, such as in Patroumpas et al. [25]. This practice may easily overcome syntactic and schematic mismatches between distinct data and metadata representations (http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/1383_abstract.pdf). Still, addressing semantic heterogeneity remains an open issue [14,26], and XSLT, albeit a Turing-complete formalism, does not constitute an optimal solution.

In fact, mapping via XSLT, a practice typically referred to as schema crosswalk, can at best prevent information loss by fine-tuning the associated style sheet, but cannot augment the information entailed by a metadata record. Instead, this can be achieved by relating individual items to more fine-grained descriptions identified by URIs. Semantic lift implicitly helps with translating metadata from a specific schema to another, also minimizing information loss with respect to schema crosswalk. As an example, if we consider translation from the ANZLIC [27] to the INSPIRE profiles of ISO metadata, it is no surprise that the former has no notion of INSPIRE Theme, a metadata item obviously mandatory in the latter. Thus, it is not possible to effectively crosswalk the two formats in this direction. Instead, semantic characterization of the keywords or topic categories that are associated with the former may

lead to the specification of INSPIRE Themes (e.g., by mapping the codelist featuring the latter with thesauri whose terms could be identified in ANZLIC metadata). Moreover, the use case we present in Section 3.1 will show that semantic lift is also functional to restore information that has been lost in the transition from an abstract metadata profile definition to its actual encoding (in our use case, in the translation from RNDT to ISO 19139). Figure 2 portrays some of the novel discovery mechanisms that are elicited by semantic enrichment of metadata. Note that the inefficiencies in discovery subtended by Use Cases (a) and (b) constitute typical examples of semantic heterogeneity.

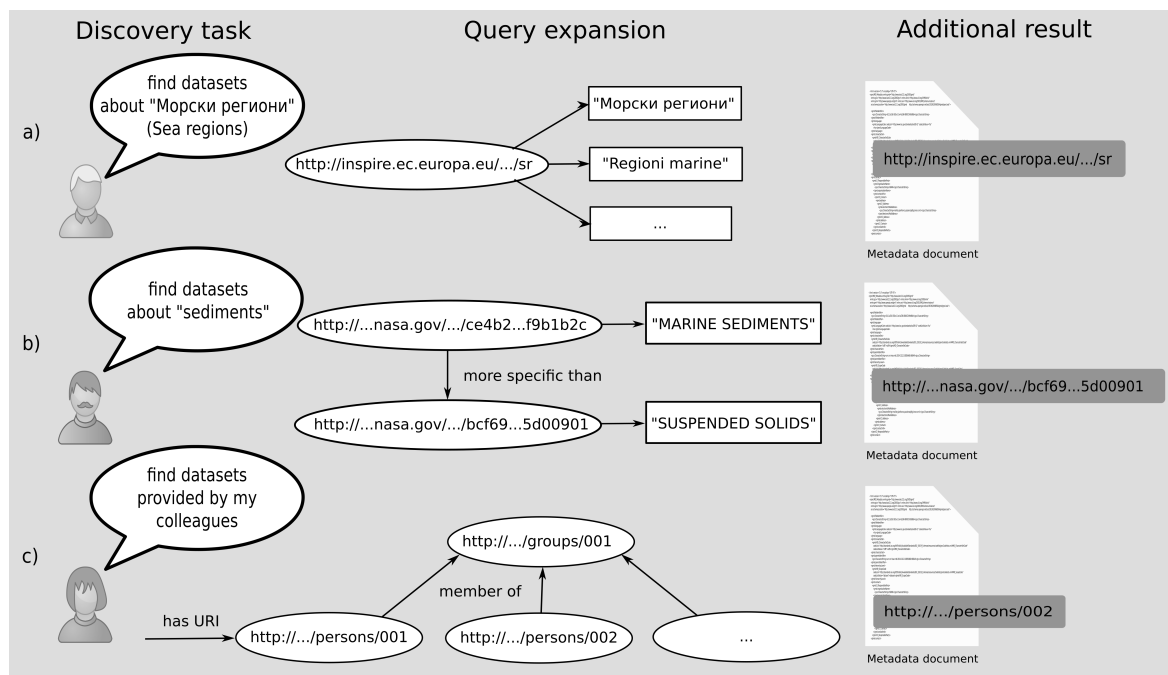


Figure 2. Use cases exemplifying semantic discovery mechanisms.

Multilingualism is the most apparent advantage of substituting text-based metadata items with unique, language-neutral identifiers (URIs in the implementation described in this paper). In Figure 2a, the user searches for "Морски региони" (the Bulgarian word for "Sea regions"): This query is likely to produce few matches, that is only datasets whose metadata are expressed in Bulgarian. Instead, by looking up the corresponding language-neutral identifier in the controlled vocabulary defined for INSPIRE Themes (<http://inspire.ec.europa.eu/theme/sr>), it is possible to expand the query in order to encompass metadata in any European language, such as the Italian counterpart "Regioni marine" in the figure. In this example, query expansion can be applied to the search term, such as in [28], independently of the underlying metadata. However, this technique can be employed to its full potential only when metadata feature this important characterization, such as in [29]. In fact, the availability of native semantic information in metadata relieves the discovery application from making multiple requests (e.g., one for each translation of the search term). This is particularly important when federating queries to multiple catalogs.

Another technique that SDIs can implement to improve recall in discovery is grounded on the organization of terms in thesauri (i.e., controlled vocabularies hierarchically articulated according to specificity). In Figure 2b, the user searches for "sediments", which is found in "MARINE SEDIMENTS", the English text representation of a term from the Science Keywords thesaurus provided by NASA through the GCMD initiative (<https://gcmd.nasa.gov/>). This term has a number of more specific terms and, among these, term "SUSPENDED SOLIDS", which is likely to be selected by metadata creators when annotating datasets on suspended matter (such as the datasets in the worked-out example of Section 5). If the metadata for these relevant resources only contain the human-intelligible

text representation associated with the latter term, “SUSPENDED SOLIDS”, they will probably not score a match for search term “sediments”. Instead, if the metadata directly refers to the term via its URI, it is possible to return these relevant resources notwithstanding the syntactic distance between the search pattern and the term’s text representation.

Finally, broadening the breadth of semantic enrichment so as to encompass individuals, work groups, projects and institutions, there is a variety of discovery mechanisms that can be grounded on the social network characterizing a given CoP. Figure 2c displays a use case where the user searches for datasets that are provided by her colleagues. This is apparently an advanced discovery functionality, and nevertheless, it can be easily achieved with our URI-based management of metadata. The only prerequisite with respect to the previous examples is authentication of the user issuing the query (in order to assign her the URI in the bottom-left part of Figure 2) and the corresponding data structures organizing people in the CoP according to work groups, projects, etc. It should be noted that it could also be feasible to transparently infer these groups on the basis of the topics of interest that users may have defined. Anyway, it is then possible to look up the individuals that share common interests and contexts with the user and then return the datasets that include these individuals among the points of contact. It is worth stressing how this kind of functionality may be important in Web 2.0 and recommendation-based discovery applications.

2.1. State of the Art

This work can be broadly ascribed to the context of geospatial semantics [30]. In particular, it relates to encoding of metadata in a semantics-aware fashion, as opposed to modeling of data and metadata by using ontologies [31]. As a consequence, important features that are in the scope of the latter, such as inference, subsumption and disambiguation, are not pertinent to this paper. Nevertheless, the expressiveness of the data structures that are associated with metadata via URIs (e.g., a SKOS thesaurus [32], rather than a full-fledged OWL ontology [33]) determine the extent of the fine-grained discovery mechanisms that can be implemented, such as query expansion, another topic not directly in the scope of this research.

The overall intent of our metadata management framework, of which Liftboy is an essential component, is to connect geospatial resources to entities in the Web of Data (resources in Semantic Web jargon), that is data structures expressed as RDF [34], identified by URIs and made accessible as either Linked Data [35] and/or (more conveniently for exploitation purposes) via SPARQL endpoints [36]. This way, referring to the distinction made by Chrisman [37], the spatial, temporal and thematic characterizations embodied by geospatial metadata can be pinpointed and disambiguated. We extended the outreach of geospatial semantics beyond these three components inasmuch as our framework can encompass any generic RDF data structure in the Web of Data, such as the individuals and organizations that are indicated as the responsible parties for a given geospatial resource. Even just a sneak peek at the vastness of the Linking Data cloud diagram (<http://lod-cloud.net/>) can suggest a number of resources that may be worth relating to metadata.

As asserted by Riedemann [38], semantic annotation is a pivotal issue in fostering interoperability among SDIs; it becomes an enabling factor in the context of web services [39] and data mediation [40]. Dill et al. [41] and Mahmoudi et al. [42] concentrate on the methodology for matching semantics-aware data structures with geospatial entities, while the focus of our work is on the identification of target metadata items in traditional metadata and the specification of appropriate SPARQL queries that can let semantic information emerge. Klien et al. [43,44] address semantic lift of geospatial data, rather than metadata, by relying on selected, authoritative ontologies; instead, de Andrade et al. [45] lift geospatial data by relying on the Web of Data at large. Our methodology endorses the latter approach, but applies to metadata instead of data. Vockner et al. [46] present a similar approach, but concentrate on cross-language information retrieval (the simpler of the use cases presented in this section). It also hints at the importance of user context (based on IP-based location and language settings), while our approach encompasses a more holistic notion of context (the third use case in Figure 2) that is derived

from user profiles. Nowak et al. [47] also elaborates on cross-language discovery, but this use case, albeit an important one, is only one of the possible applications enabled by semantics-aware metadata.

In [48], Li et al. exploit LSA to improve geospatial discovery by applying this human-independent method to metadata indexing. In fact, the paper acknowledges that methodologies grounded on selected ontologies and, in general, semantic data sources for either indexing of resources or query expansion (such as in [49,50]) are often biased by subjectivity in the organization of these sources. This is particularly relevant in contexts, such as that of geosciences, where distinct CoP typically employ different terminologies. In our work, this is the main reason for allowing the system administrator to customize the semantic sources that are employed for semantic lift (as explained in the following section). Moreover, albeit that this technique allows for better results with respect to traditional full-text keyword matching, in our opinion, it falls short of providing reusable mappings to semantic resources, which instead are made available by our application as WADM annotations.

Adeyinka et al. [51] address subjectivity in the selection and organization of terms by employing upper-level characterizations (e.g., categorizing semantic relations among terms in conjunction with terms themselves) and applying reasoning. Still, the proposed model leverages the GeoNames ontology (<http://www.geonames.org/>) for geographic characterization, and this may lead to the same heterogeneities it aims to avoid. As an example, in encoding as RDF the LifeWatch Italy toponyms (<http://fuseki1.get-it.it/LWItaToponyms/>) [52], which are based on the authoritative source for Italian toponyms, and mapping them to GeoNames, we spotted a number of mismatches that are prone to emerge in the mapping of any national categorization of toponyms with GeoNames. Furthermore, our aim is to support the creation of semantic indices as the seminal task that may lead to novel discovery paradigms rather than improving existing ones. The use cases presented in this section stick to a query-response paradigm because, for the time being, our research is not mature enough to propose new ones.

3. EDI and Liftboy for Semantic Lift

The EDImetadata editor [17] marked a breakthrough in assisted editing of metadata descriptions in the geospatial domain. In fact, besides being conceived of in the context of the Italian flagship project RITMARE (<http://ritmare.it>), EDI has been adopted so far by a number of projects, namely the FP7 projects ERMES (<http://www.ermes-fp7space.eu/it/homepage/>) and EuroFleets2 (<http://www.eurofleets.eu/np4/home.html>), the H2020 project eLTER (<http://www.lter-europe.net/lter-europe/projects/eLTER>) and the Italian Flagship Project NextData (<http://www.nextdataport.it/>). The works in [21,53] describe the workflow by means of which semantically-enriched metadata are created from scratch using EDI in an ex-ante approach; this practice corresponds to the upper part of Figure 3. Still, there is large amount of metadata that did not take advantage of the tool at editing-time and thus lacks any semantic characterization. The application we present in this paper, Liftboy, reconstructs the missing semantic information that EDI creates as a transparent by-product of metadata editing. On the one hand, this practice elicits the query expansion functionalities exemplified in the previous section. On the other, once complemented with the appropriate URIs, metadata descriptions can be updated via EDI as if originally created with this tool. The system administrators who regulate the behavior of the EDI editor can customize the behavior of Liftboy by leveraging the same data structures and the same template language. In fact, as explained in the following paragraphs of this section, all information that is necessary for the semantic lift task is already featured in template definitions. Hence, any alternative data structure would be largely isomorphic to the template language that has been defined. Moreover, in order to enable re-editing of lifted metadata through EDI, the output of either the latter and that of Liftboy shall be the same. Figure 3 depicts both the workflow to create enriched metadata by using EDI (the upper part) and the workflow for ingestion of traditional metadata in our management framework through Liftboy (the lower part). The description of the former is in [53], while we detail here the second one. The tailoring of an appropriate template (1) is still a necessary preliminary activity and even a more important one because the

system administrator is required to produce two templates instead of just one. In fact, it is unlikely that the template used for lifting metadata (indicated as “Liftboy template” in Figure 3) can also be subsequently used by the metadata maintainer to edit metadata with EDI, as the former is typically a subset of the template that is aimed at the human agent (the rationale for this is to be explained in the next section). As an example, Listing 1 portrays the key definitions in the template for ISO 19135 metadata expressing the e-mail of the responsible party for a dataset. Please refer to [53] for a thorough explanation of template constructs.

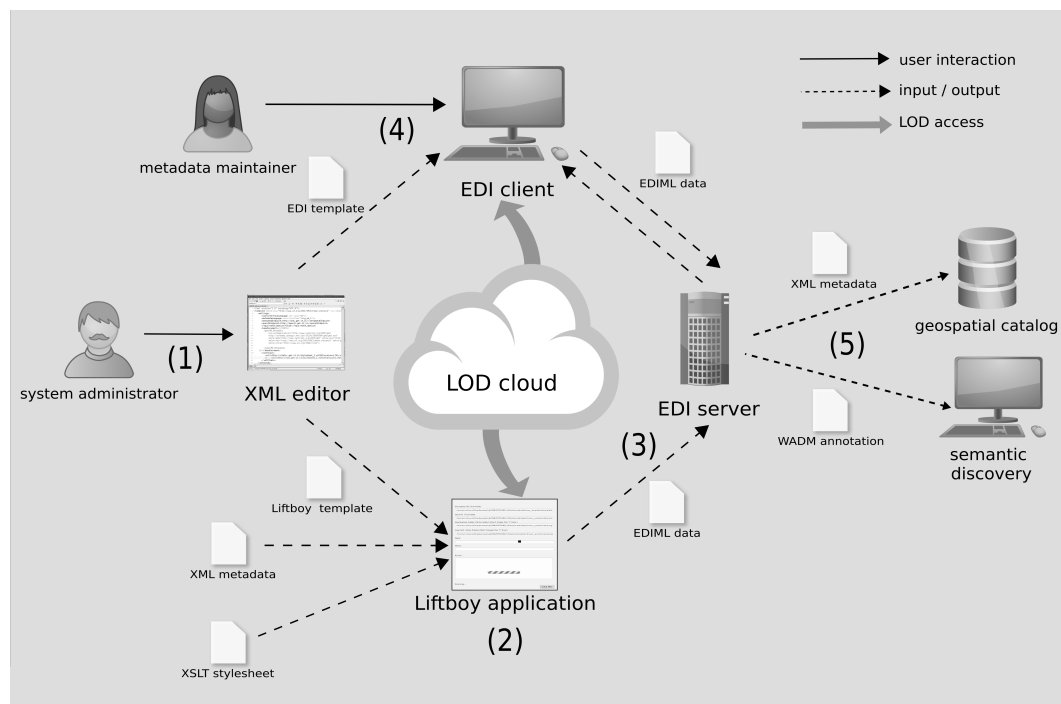


Figure 3. Overview of workflows for semantic lift through EDI/Liftboy.

Listing 1: Code fragment from an EDI/Liftboy template defining the essentials on a metadata item

```

1 <template>
2 <sparql xml:id="person">
3 <query><![CDATA[
4   SELECT ?contact ?label
5   WHERE {
6     ?contact rdf:type foaf:Person .
7     ?contact vcard:email ?label .
8     FILTER( REGEX( STR(?label), "\$search_param", "i" ) ) }
9   ORDER BY ASC(?label)
10 ]]></query>
11 <url>http://some.endpoint.org/sparql</url>
12 </sparql>
13 <element xml:id="resp">
14 <label xml:lang="en">Responsible party</label>
15 <hasRoot>/gmd:MD_Metadata/.../gmd:CI_Citation</hasRoot>
16 ...
17 <produces>
18 <item
19   hasDatatype="autoCompletion" datasource="person">
20   <label xml:lang="en">Email</label>
21   <hasPath>

```

```

22      gmd:citedResponsibleParty/.../gmd:electronicMailAddress/...
23      </hasPath>
24      ...
25    </item>
26    ...
27  </produces>
28 </element>
29 <template>

```

The intuition behind the development of Liftboy is the following: Referring to the listing, the two core definitions for the metadata item are:

1. The specification of the XML element to be produced for the specific metadata item, expressed as the XPath expression obtained by combining the paths in Lines 15 and 22. Note that the hasRoot tag is functional to indicate the root element where multiple instances of the same metadata element shall be nested.
2. The specification of data source “person” (Line 2) as the one governing autocompletion of the item (Lines 18–25), on the basis of the output of the SPARQL query defined in Lines 4–9. Note that the client-side component of EDI substitutes placeholder “\$search_param” in Line 8 with the characters entered by the user in the corresponding form field of the interface.

It is apparent that this information can be used “the other way around” for locating the text-based metadata item in input ISO 19136 metadata and look up the data source to check whether the entity represented by the item (in this case, an individual) has a semantic counterpart in the triple store underlying the SPARQL endpoint that is defined. As an example, placeholder “\$search_param” is going to be substituted with the text contained in the XML element identified by the XPath in 1.

The Liftboy template is then fed to Liftboy together with the input metadata that require semantic augmentation (2); the template also specifies the XSLT style sheet that the application can use to tailor the output to the intended EDI template. The output format is EDIML, the internal XML storage format hinted at above, that is stored by the EDI server (3). Harnessing these data, EDI can allow the metadata maintainer to update the description retaining the reconstructed semantic information (4). The last phase in the workflow (5) consists of the generation of the appropriate XML metadata description (say, ISO 19139) that can be inserted into a traditional geospatial catalog, as well as a WADM annotation [22] containing the semantic information that can be used to ground the fine-grained discovery mechanisms exemplified in Figure 2. Serialization of semantic information as WADM annotations is detailed in Section 4.

Figure 4 portrays the interface to the Liftboy application, which has been kept as simple as possible because; as will become clear in the following paragraphs of this section, most logics reside in the template that is processed by Liftboy, as well as in the selection of RDF data sources that are specified in the latter. Hence, the only information required for functioning is the specification of an appropriate template (Figure 4(1)), the source metadata the template shall be applied to (Figure 4(2)) and the output directory where the generated EDIML file shall be created, together with a process log (Figure 4(3)). The interface also provides a text box for displaying error messages. The remaining piece of information that is necessary for the enactment of the workflow in Figure 3, that is the XSLT style sheet that makes the generated EDIML apt at re-editing via EDI, can be directly specified in the template.

Figure 4. Interface to the Liftboy application.

3.1. Writing Templates for Liftboy

The reason why the template that is fed to Liftboy, Phase (2) in the workflow, may not be the template to be used later on when updating the metadata in Phase (4) is two-fold. On the one hand, the system administrator can benefit from a far greater flexibility in articulating Liftboy templates w.r.t. those aimed at EDI. On the other hand, Liftboy templates are applied to metadata descriptions that may suffer from information loss due to the XML encoding in the target metadata schema. These important differences are detailed in this section.

3.1.1. Flexibility in Liftboy Templates

By considering the role of XPath definitions in EDI templates [53], it is apparent that the only common requirement with Liftboy templates is that they need to be absolute location paths. In fact, since XPath expressions in EDI templates specify the XML element or attribute to be created and populated with user input, from the document root to leaf node, they can only use the unabbreviated syntax [54]. Otherwise, by using axes (such as “ancestor::”) or shortcuts (such as “//” and “..”), it is possible to violate this requirement. Moreover, these expressions may not contain predicates, that is conditions in square brackets such as “[position() > 1]”, as these may identify node sets instead of just a single node.

On the contrary, Liftboy templates can benefit from the full expressiveness of XPath because expressions are just meant to locate one or more nodes whose content can be related to URIs. As an example, in order to pinpoint the e-mails of points of contact whose institution is “http://acme.org”, one could insert in the template an XPath expression such as the following:

```
//gmd:electronicMailAddress[ancestor::gmd:CI_Contact//gmd:URL='http://acme.org']
```

3.1.2. Information Loss in Metadata Encoding

Unless the encoding of a given metadata format is specifically tied to the intended profile, it is unlikely that bidirectional translation from the actual encoding to the abstract profile is possible. This is apparent in many profiles based on ISO 19115/19119, such as RNDT, the metadata profile mandated by the transposition of the INSPIRE Directive [4] in Italian law. As an example, RNDT mandates three different categories of points of contact for the dataset (“Responsible party”, “Point of contact” and “Distributor”), but upon serialization of metadata in the ISO 19135 format, the three categories generate the same XML elements and can be distinguished only by considering the role that is associated with each individual.

Editing of RNDT metadata with EDI (http://edidemo.get-it.it/dist/RNDT_dataset.html) allows for discriminating between the three different categories of point of contact defined by the profile, but when re-editing metadata that has been processed by Liftboy, this distinction is prone to disappear. Of course, it is possible to create specific template structures for each of the aforementioned categories (akin to those in Listing 1), but the semantic lift is going to be error prone (e.g., the template shall distinguish between the Italian and English representations of roles that, in ISO 19136 encoding, are plain text). A better solution is to apply the optional XSLT style sheet indicated in Figure 3 to expand the EDIML output produced in Phase (3) so as to reflect the EDI template that is employed for updating the metadata record.

4. Decoupling Semantic Information from Metadata via WADM Annotations

In the context of Linked Data, WADM [22] provides a standardized way to represent associations between distinct web resources. This paradigm perfectly fits our notion of metadata as distributed representations. In particular, the semantic lift operated by Liftboy can be regarded as a specific kind of document annotation, relating fragments of a metadata document (such as the ISO 19139 records considered in this work) to semantic resources (i.e., RDF data structures identified by URIs). More specifically, semantic lift can be regarded as a mapping

$$S : \{MetadataDocuments\} \rightarrow \{AugmentedMetadataDocuments\}$$

An AugmentedMetadataDocument is intended, following the definitions in [55], as a document with pointers (dereferenceable URIs) to semantic information defining the precise nature of some item. In this section, we sketch the WADM encoding of the Liftboy output presented above as a general, standardized way to convey the semantic information entailed by metadata. Listing 2 portrays the fragment of the EDIML code generated by Liftboy, which corresponds to template definitions in Listing 1. The key components that are necessary to serializing this information as WADM annotations are:

1. The specification of the source metadata file that is the subject of the annotations (Line 3).
2. The specific path pinpointing the metadata element under consideration (Lines 12 and 14).
3. The URI that is related to the metadata item by semantic augmentation (Line 18).

Listing 2: Code fragment from an EDIML file generated by Liftboy

```

1 <elements>
2   ...
3   <fileUri>http://.../metadata/Dataset_ABC</fileUri>
4   ...
5   <element>
6     <id>resp</id>
7     ...
8     <items>
9       <item>
10        <id>resp_1</id>
11        <element_id>resp</element_id>
12        <hasRoot>/gmd:MD_Metadata/.../gmd:CI_Citation</hasRoot>
13        <path>
14          gmd:citedResponsibleParty/.../gmd:electronicMailAddress/...
15        </path>
16        <value>john.doe@acme.org</value>
17        <codeValue>
18          http://acme.org/personnel/JohnDoe
19        </codeValue>
20        ...

```

```

21     </item>
22     ...
23     <items>
24     ...
25 </element>
26 <elements>

```

Listing 3 presents the semantic lift as a WADM annotation serialized as JSON-LD (<https://json-ld.org/spec/latest/>), an extension to JSON that allows one to express RDF triples as JSON key-value pairs. In WADM terminology, the item in the source metadata document is the target of the annotation (Lines 21–28), while the referenced semantic resource is the body of the annotation (Lines 17–20). Going into detail, both the target and the body are composed of an id that is their (dereferenceable) URI and by a format attribute, specifying the expected media-type of the resources: In the example, the ISO 19139 media type (e.g., INSPIRE metadata) is specified as the target format, and RDF + XML (the XML serialization of RDF) as the body format. The structure specified by the selector attribute of the target serves the purpose of selecting a specific item in the target, and in this case, it is specified that the selection must be performed through the XPath expression (XPathSelector) defined in the value attribute, that in our scenario corresponds to the concatenation of the hasRoot and path elements in the EDIML file.

Listing 3: Web Annotation derived from the EDIML generated by Liftboy

```

1  {
2    "@context": "http://www.w3.org/ns/anno.jsonld",
3    "id": "http://example.org/anno1",
4    "type": "Annotation",
5    "motivation": "identifying",
6    "creator": {
7      "id": "http://example.org/liftboy/",
8      "type": "Software",
9      "name": "Liftboy v1.0beta",
10     "nickname": "Liftboy",
11     "homepage": "https://github.com/SP7-Ritmare/Liftboy/"
12   },
13   "created": "2015-01-28T12:00:00Z",
14   "modified": "2015-01-29T09:00:00Z",
15   "generator": "http://example.org/liftboy",
16   "generated": "2015-02-04T12:00:00Z",
17   "body": {
18     "id": "http://acme.org/personnel/JohnDoe",
19     "format": "rdf+xml"
20   },
21   "target": {
22     "id": "http://.../metadata/Dataset_ABC",
23     "format": "vnd.iso.19139+xml",
24     "selector": {
25       "type": "XPathSelector",
26       "value": "/gmd:MD_Metadata/.../gmd:CI_Citation/
27       gmd:citedResponsibleParty/.../gmd:electronicMailAddress/..."
28     }
29   }
30 }

```

In the example WADM annotation, among other attributes that are self-explanatory, the choice of value “identifying” for key motivation - the code to be used when the user intends to assign an identity to the target (<http://www.w3.org/ns/oa#identifying>) - deserves some more words. In WADM, several motivation values are proposed, defined in [56]. For the present case, the identification of some item with a specific RDF resource seemed to fit at best. On the other hand, maybe it could be of some interest to define “semantic lift” as a narrower concept of “identifying”, following the extensible design of motivations suggested in Annex C of [56].

Finally, Listing 4, groups in an annotation collection all the annotations related to each semantic lift produced by the software for the original metadata document. In the example, the collection counts two (Line 13) annotations as items (Lines 19 and 20) of an embedded page (for a discussion of annotation collection pages, see [22]).

Listing 4: Annotation Collection with embedded annotation page

```

1  {
2    "@context": "http://www.w3.org/ns/anno.jsonld",
3    "id": "http://example.org/collection1",
4    "type": "AnnotationCollection",
5    "label": "Semantic annotation collection",
6    "creator": {
7      "id": "http://example.org/user1",
8      "type": "Software",
9      "name": "Liftboy v1.0beta",
10     "nickname": "Liftboy",
11     "homepage": "https://github.com/SP7-Ritmare/Liftboy/"
12   },
13   "total": 2,
14   "first": {
15     "id": "http://example.org/page1",
16     "type": "AnnotationPage",
17     "startIndex": 0,
18     "items": [
19       "http://example.org/anno1",
20       "http://example.org/anno2"
21     ]
22   }
23 }
```

Appendices A–D provide an example annotation walkthrough with respect to another essential component of metadata, that is keywords expressing the thematic context of a geospatial dataset.

5. Exploiting Semantic Information

In this section, we revisit the use cases presented in Section 2 and show how semantic enrichment of metadata achieves the intended goals. These worked-out examples were carried out on a repository hosted by the CNR IREA institute in Milan (<http://skmi.irea.cnr.it/>) and containing the map of TSM (Total Suspended Matter) obtained from imagery data acquired from MERIS on the ESA Envisat satellite. A subset of the datasets has been annotated according to the RNDT profile of INSPIRE metadata considered so far and can be browsed or discovered by means of the search functionalities provided by the geoportal. Then, the ISO 19139 metadata have been downloaded, lifted with the application presented in this work, translated into RDF (the XSLT style-sheet performing this transformation omitted for brevity) and uploaded to a SPARQL endpoint (<http://fuseki3.get-it.it/dataset.html?tab=query&ds=/LiftboyData>). The queries presented in the

following of this section can be reproduced in either the discovery interface of the geoportal or in the query form of the SPARQL endpoint.

5.1. Cross-Language Discovery

Searching for “Regioni marine” by means of the search interface provided by the geoportal (a front-end to the underlying CSW implementation written in Python, <http://pycsw.org/>), the user can retrieve the datasets that were associated the Italian translation of this INSPIRE Theme. The user is going to retrieve an empty result set when searching for the same theme in any other language (except, of course, in the case of apparent syntactic similarity), such as “Морски региони”, the corresponding translation in Bulgarian. Instead, by issuing the query in Listing 5 to the SPARQL endpoint, the whole set of results produced by the geoportal is returned:

Listing 5: SPARQL query implementing cross-language discovery

```

1 PREFIX def:      <http://sp7.irea.cnr.it/rdfdata/schemas#>
2 PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs:     <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX foaf:     <http://xmlns.com/foaf/0.1/>
5 PREFIX xsd:      <http://www.w3.org/2001/XMLSchema#>
6 PREFIX sp7:      <http://sp7.irea.cnr.it/rdfdata/project/>
7 PREFIX skos:     <http://www.w3.org/2004/02/skos/core#>
8 PREFIX dcat:     <http://www.w3.org/ns/dcat#>
9 PREFIX dct:      <http://purl.org/dc/terms/>
10
11 SELECT DISTINCT ?dataset ?title
12 WHERE {
13     ?dataset      rdf:type          dcat:Dataset .
14     ?dataset      dct:title         ?title .
15     ?dataset      dcat:theme        ?keyword .
16     ?keyword      skos:prefLabel    ?label .
17     FILTER( REGEX( STR(?label), "Морски региони", "i") )
18 }
```

The query is straightforward, matching the text representations of keywords with the search pattern (Lines 15–17), except that the search pattern can be matched against a broader range of language-dependent representations. Obviously, the breadth of the cross-language capabilities that are elicited by our RDF-based representation of metadata is dependent on the multilingual thesauri that are plugged into Liftboy templates. The RDF metadata for any of the results can be retrieved by issuing a DESCRIBE query such as the following:

```
DESCRIBE <http://sp7.irea.cnr.it/data/6c43fd22-9659-11e4-a8cd-5254007ad55c>
```

5.2. Query Expansion

Searching for “sediments” by the geoportal’s catalog is not going to return any match because this pattern cannot be found in the XML-based metadata records. Nevertheless, this term is a viable replacement for “suspended matter” because the latter is a specific category of sediments, at least according to GCMD. This thesaurus has been referred to in the template used for semantic lift of resource metadata, and then, the SPARQL query in Listing 6 is going to return the same results of a query for “suspended matter” by the geoportal’s catalog.

Lines 16–18 retrieve the terms that are broader (i.e., more general) than the keywords specified in metadata and match the search pattern against their textual representation (SKOS `prefLabels`). Note that, for ease of enactment of the query, the RDF dump of the GCMD Science Keywords thesaurus has been downloaded and duplicated in the triple store. However, SPARQL 1.1-compliant endpoints

allow for federated queries via the SERVICE clause; it is then possible to encompass in the discovery process all the external endpoints that may have been specified in the template.

Listing 6: Query implementing query expansion

```

1  PREFIX def:      <http://sp7.irea.cnr.it/rdfdata/schemas#>
2  PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX rdfs:     <http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX foaf:     <http://xmlns.com/foaf/0.1/>
5  PREFIX xsd:      <http://www.w3.org/2001/XMLSchema#>
6  PREFIX sp7:      <http://sp7.irea.cnr.it/rdfdata/project/>
7  PREFIX skos:     <http://www.w3.org/2004/02/skos/core#>
8  PREFIX dcat:     <http://www.w3.org/ns/dcat#>
9  PREFIX dct:      <http://purl.org/dc/terms/>
10
11 SELECT DISTINCT ?dataset ?title
12 WHERE {
13     ?dataset          rdf:type          dcat:Dataset .
14     ?dataset          dct:title         ?title .
15     ?dataset          dcat:theme        ?keyword .
16     ?keyword          skos:broader      ?broader_term .
17     ?broader_term     skos:prefLabel   ?label .
18     FILTER( REGEX( STR(?label), "sediments", "i") )
19 }
```

5.3. Expansion Based on Social Network

The third use case presented in Section 2 has no counterpart in traditional discovery and, thus, cannot be performed in the search interface provided by the geoportal. In the query in Listing 7, the URI identifying the user in Figure 2c is actualized with that of one of the authors, sp7:CristianoFugazzaIREA (the preamble sp7: standing for the longer path in Line 6). The query retrieves the datasets that have, among the points of contact defined in the metadata, an individual from the same organization. Of course, the query can be modified in order to match individuals in an arbitrary social network (e.g., matching users in the same foaf:Group, users that are linked by foaf:knows properties), etc.

Listing 7: Query expansion based on social network

```

1  PREFIX def:      <http://sp7.irea.cnr.it/rdfdata/schemas#>
2  PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX rdfs:     <http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX foaf:     <http://xmlns.com/foaf/0.1/>
5  PREFIX xsd:      <http://www.w3.org/2001/XMLSchema#>
6  PREFIX sp7:      <http://sp7.irea.cnr.it/rdfdata/project/>
7  PREFIX skos:     <http://www.w3.org/2004/02/skos/core#>
8  PREFIX dcat:     <http://www.w3.org/ns/dcat#>
9  PREFIX dct:      <http://purl.org/dc/terms/>
10 PREFIX vcard:    <http://www.w3.org/2006/vcard/ns#>
11
12 SELECT DISTINCT ?dataset ?title
13 WHERE {
14     sp7:CristianoFugazzaIREA      vcard:org      ?organization .
15     ?dataset          rdf:type          dcat:Dataset .
16     ?dataset          dct:title         ?title .
17     ?dataset          dcat:contactPoint/vcard:hasUID ?person .

```

```

18         ?person          vcard:org          ?organization .
19     }

```

For the sake of clarity, the queries in this section implement the specific use cases presented in Section 2. However, it is apparent that it is possible to conceive of a single query that applies a wider spectrum of semantics-aware functionalities to the metadata of datasets.

6. Conclusions

The metadata management framework we developed for project RITMARE aims at enriching metadata descriptions with semantic information for the purpose of enabling novel discovery mechanisms. In our framework, metadata records created from scratch already have this important characteristic, but in order to ingest pre-existing metadata corpora, it was necessary to execute semantic lift on generic resource descriptions. This is achieved by means of the Liftboy application, a Java FOSS that repurposes the template-driven philosophy already implemented in the metadata editing facilities we developed. Liftboy allows for seamless integration of existing resource descriptions in our metadata management workflow. We tested the application in the context of ISO 19115/19119 metadata, but the methodology is applicable to any XML-based metadata schema. Moreover, the application can harness generic SPARQL endpoints as the data sources that are considered for the semantic lift task. Finally, the output of this process is provided as WADM annotations in order to single out the semantic information in a metadata record for use by discovery applications.

In this paper, we provided the rationale for semantic lift in geospatial resource management as it fosters normalized, multi-tenanted articulation of metadata descriptions. As introduced in Section 1 and detailed in [21], referring to the authoritative source for a given piece of information via its URI allows applications to generate a metadata record on demand (i.e., when the user requests it for download) and, thus, to access an up-to-date version of the metadata item. This practice elicits dramatic improvements with regard to the consistency of metadata descriptions. Liftboy also has a more extensive applicability to geospatial metadata because, albeit being aimed at inclusion of pre-existing resource descriptions in the RITMARE framework, it also allows for semantic indexing of third-party datasets. In fact, the provision of mappings to RDF data structures as WADM annotations realizes decoupling of semantic information from the originating metadata. Thus, it is straightforward to implement fine-grained discovery mechanisms on resources that are brokered from external catalogs, such as in [57,58].

Moreover, further opportunities arise when considering serialization of semantic lift information as WADM annotations in the perspective of LDP [59] and WAP [60] applications. As an example, consider the scenario where the semantic lift is carried out by Liftboy or another agent (either human or automated) externally from the authority governing metadata; for instance, semantic lift could be performed during metadata harvesting by an external discovery service. The external service could then expose an augmented version of the original metadata document, with no possibility to update the original one. At the same time, it would be useful for the originating source acknowledge the process and the augmented version, in order to be able to update its own metadata accordingly. It could perhaps be the case that the original authority would not accept certain parts of the augmentation obtained by the external process, for example because of a different interpretation of some semantic lift result (e.g., the authority could be aware of a more specific concept to associate with a metadata item). Such flexibility is built (to our knowledge, only) in the framework proposed for WADM annotations.

Although the development of Liftboy is stable, we are prone to consider the software as still prototypical because of the sensitivity of the output to the template that is fed to the application, as well as to the RDF data sources that are accessed for performing semantic lift. Furthermore, it is not straightforward to interpret the gain in performance in terms of the typical measurable indices for IR. This is the main reason why, in Section 5, we provided qualitative results rather than quantitative ones. The release of the production-grade version of EDI (<http://edidemo.get-it.it/>) predated finalization of the

templates that implement the broad range of metadata schemas that are currently made available in the distribution. We expect a similar delay in the development of effective Liftboy templates. In fact, recourse to the metalanguage templates is based on shift fine tuning of the application from the programmer to the domain expert, but cannot contract development time significantly.

Liftboy is currently available as a desktop application in order to allow for easy testing and rapid development of templates. However, Liftboy is ideally meant for integration with the GET-IT software suite (<http://get-it.it/>), our comprehensive geoportal software based on GeoNode. In fact, GET-IT currently cannot ingest pre-existing metadata descriptions: Liftboy is going to entail migration of traditional catalogs to a novel, semantics-aware inflection of geospatial information.

Acknowledgments: The activities described in this paper have been funded by project RITMARE. We also acknowledge the support of LifeWatch Italy.

Author Contributions: Cristiano Fugazza conceived of and designed the application. Luca Frigerio developed Liftboy. Paola Carrara coordinated the research group. Cristiano Fugazza, Paolo Tagliolato and Paola Carrara wrote the paper. All authors read and approved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANZLIC	Australia New Zealand Land Information Council
CoP	Community of Practice
CNR-IREA	National Research Council, Institute for Electromagnetic Sensing of the Environment
EMF	Environmental Monitoring Facilities
ENVRI	ENVironmental Research Infrastructures
ERMES	Earth obseRvation Model based RicE information Service
EU	European Union
FOSS	Free and Open Source Software
GCMD	Global Change Master Directory
GEOSS	GEO System of Systems
GET-IT	Geoinformation Enabling ToolKIT
GIS	Geographic Information Systems
INSPIRE	INfrastructure for Spatial InfoRmation in Europe
IR	Information Retrieval
IT	Information Technology
JSON-LD	JavaScript Object Notation for Linking Data
LDP	Linked Data Platform
LSA	Latent Semantic Analysis
LTER	Long-Term Ecosystem Research
OGC	Open Geospatial Consortium
RDF	Resource Description Framework
RNDT	Repertorio Nazionale dei Dati Territoriali (National Repository of Territorial Data)
RITMARE	Ricerca ITaliana per il MARE (Italian Research for the Sea)
SDI	Spatial Data Infrastructures
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier
XPath	XML Path Language
XSLT	eXtensible Stylesheet Language Transformations
WADM	Web Annotation Data Model
WAP	Web Annotation Protocol

Appendix A. EDI/Liftboy Template Fragment Defining a Keyword in INSPIRE Metadata

```

1 <element xml:id="keyw_voc_contr" isMandatory="false" isMultiple="true">
2   <label xml:lang="en">Keyword from controlled vocabularies</label>
3   <hasRoot>/gmd:MD_Metadata/.../gmd:MD_DataIdentification</hasRoot>
4   <produces>
5     <item hasDatatype="autoCompletion" datasource="keywordFromContrVoc">
6       <label xml:lang="en">Keyword</label>
7       <hasPath>
8         gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyword/...
9       </hasPath>
10    </item>
11    ...
12  </produces>
13 </element>

```

Appendix B. Fragment from ISO 19139 Metadata Defining a Keyword

```

1 <gmd:descriptiveKeywords>
2   <gmd:MD_Keywords>
3     <gmd:keyword>
4       <gco:CharacterString>Ocean Currents</gco:CharacterString>
5     </gmd:keyword>
6   </gmd:MD_Keywords>
7   ...
8 </gmd:descriptiveKeywords>

```

Appendix C. EDIML Produced by Liftboy Relative to the Keyword in Appendix B

```

1 <element>
2   <id>keyw_voc_contr</id>
3   ...
4   <items>
5     <item>
6       <id>keyw_voc_contr_1</id>
7       <element_Id>keyw_voc_contr</element_Id>
8       <path>
9         /gmd:MD_Metadata/.../gmd:descriptiveKeywords/.../gmd:keyword/...
10      </path>
11      <value>Ocean Currents</value>
12      <codeValue>http://gcmd.gsfc.nasa.gov/skos#ocean_currents</codeValue>
13      ...
14    </item>
15    ...
16  </items>
17 </element>

```

Appendix D. WADM Annotation Relative to the Keyword Item in Appendix B

```

1 {
2   "@context": "http://www.w3.org/ns/anno.jsonld",
3   "id": "http://example.org/anno2",
4   "type": "Annotation",
5   "motivation": "identifying",

```

```

6   ...
7   "body": {
8     "id": "http://gcmd.gsfc.nasa.gov/skos#ocean_currents",
9     "format": "rdf+xml"
10  },
11  "target": {
12    "id": "http://.../metadata/Dataset_ABC",
13    "format": "vnd.iso.19139+xml",
14    "selector": {
15      "type": "XPathSelector",
16      "value": "/gmd:MD_Metadata/.../gmd:descriptiveKeywords/.../gmd:keyword/
17      gco:CharacterString[text()='Ocean Currents']"
18    }
19  }
20 }

```

References

- Salvemini, M. From the GIS to the SDI: A design path. In Proceedings of the 7th AGILE Conference on Geographic Information Science, Crete, Greece, 29 April–1 May 2004; pp. 1–7.
- Craglia, M.; Nativi, S.; Santoro, M.; Vaccari, L.; Fugazza, C. Inter-disciplinary Interoperability for Global Sustainability Research. In Proceedings of the 4th International Conference on GeoSpatial Semantics (GeoS'11), Brest, France, 12–13 May 2011; Springer-Verlag: Berlin/Heidelberg, Germany, 2011; pp. 1–15.
- Perego, A.; Fugazza, C.; Vaccari, L.; Lutz, M.; Smits, P.; Kanellopoulos, I.; Schade, S. Harmonization and Interoperability of EU Environmental Information and Services. *IEEE Intell. Syst.* **2012**, *27*, 33–39.
- European Commission. 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). 2007. Available online: <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002> (accessed on 5 November 2017).
- European Commission. 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. 2008. Available online: <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32008R1205> (accessed on 5 November 2017).
- European Commission. Corrigendum to 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. 2008. Available online: <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32008R1205R%2802%29> (accessed on 5 November 2017).
- Christian, E.J. GEOSS Architecture Principles and the GEOSS Clearinghouse. *IEEE Syst. J.* **2008**, *2*, 333–337.
- Butterfield, M.L.; Pearlman, J.S.; Vickroy, S.C. A System-of-Systems Engineering GEOSS: Architectural Approach. *IEEE Syst. J.* **2008**, *2*, 321–332.
- Gore, A. The digital earth. *Aust. Surv.* **1998**, *43*, 89–91.
- Craglia, M.; de Bie, K.; Jackson, D.; Pesaresi, M.; Remetey-Fülöpp, G.; Wang, C.; Annoni, A.; Bian, L.; Campbell, F.; Ehlers, M.; et al. Digital Earth 2020: Towards the vision for the next decade. *Int. J. Digit. Earth* **2012**, *5*, 4–21.
- Nieva de la Hidalgo, A.; Hardisty, A. *How the ENVRI Reference Model Helps to Design Research Infrastructures*; Technical Report; ENVRIplus Newsletter: 2016. Available online: www.envriplus.eu/wp-content/uploads/2016/05/ENVRI-Reference-Model.pdf (accessed on 5 November 2017).
- Bishr, Y. Overcoming the semantic and other barriers to GIS interoperability. *Int. J. Geogr. Inf. Sci.* **1998**, *12*, 299–314.
- Ouksel, A.M.; Sheth, A. Semantic Interoperability in Global Information Systems. *SIGMOD Rec.* **1999**, *28*, 5–12.
- Ramakrishnan, C.; Sheth, A.; Thomas, C. Semantics for The Semantic Web: The Implicit, the Formal and the Powerful. *Int. J. Semant. Web Inf. Syst.* **2005**, *1*, 1–18.

15. Nativi, S.; Bigagli, L. Discovery, Mediation, and Access Services for Earth Observation Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 233–240.
16. Berners-Lee, T.; Fielding, R.; Masinter, L. Uniform Resource Identifiers (URI): Generic Syntax. 1998. Available online: <https://tools.ietf.org/html/rfc3986> (accessed on 5 November 2017).
17. Pavesi, F.; Basoni, A.; Fugazza, C.; Menegon, S.; Oggioni, A.; Pepe, M.; Tagliolato, P.; Carrara, P. EDI—A Template-Driven Metadata Editor for Research Data. *J. Open Res. Softw.* **2016**, *4*, doi:10.5334/jors.106.
18. ISO. ISO 19115:2014 Geographic Information—Metadata. Standard, International Organization for Standardization (TC 211). 2014. Available online: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=53798 (accessed on 5 November 2017).
19. ISO. ISO 19119:2005 Geographic Information—Services. Standard, International Organization for Standardization (TC 211). 2005. Available online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=39890 (accessed on 5 November 2017).
20. ISO. ISO 191136:2007 Geographic Information—Geography Markup Language (GML). Standard, International Organization for Standardization (TC 211). 2007. Available online: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32554 (accessed on 5 November 2017).
21. Fugazza, C.; Pepe, M.; Oggioni, A.; Tagliolato, P.; Carrara, P. Streamlining geospatial metadata in the Semantic Web. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2016; Volume 34.
22. W3C—Web Annotation Working Group. Web Annotation Data Model. Available online: <https://www.w3.org/TR/annotation-model/> (accessed on 5 November 2017).
23. Agenzia per l'Italia Digitale (AgID). *Repertorio Nazionale dei Dati Territoriali—RNDT*; Technical Report; Agenzia per l'Italia Digitale (AgID): 2012. Rome, IT. Available online: <http://archivio.digitpa.gov.it/repertorio-nazionale-dei-dati-territoriali-rndt> (accessed on 5 November 2017).
24. W3C. XSL Transformations (XSLT) Version 1.0. 1999. Available online: <https://www.w3.org/TR/xslt> (accessed on 5 November 2017).
25. Patroumpas, K.; Georgomanolis, N.; Stratiotis, T.; Alexakis, M.; Athanasiou, S. Exposing INSPIRE on the Semantic Web. *Web Semant.* **2015**, *35*, 53–62.
26. Lutz, M.; Sprado, J.; Klien, E.; Schubert, C.; Christ, I. Overcoming semantic heterogeneity in spatial data infrastructures. *Comput. Geosci.* **2009**, *35*, 739–752.
27. Australia New Zealand Land Information Council. ANZLIC Metadata Profile. 2007. Available online: <http://www.anzlic.gov.au/resources/metadata> (accessed on 5 November 2017).
28. Santoro, M.; Mazzetti, P.; Nativi, S.; Fugazza, C.; Granell, C.; Díaz, L. Methodologies for augmented discovery of geospatial resources. In *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications*; Díaz, L., Granell, C., Huerta, J., Eds.; IGI Global: Hershey, PA, USA, 2012; Chapter 9, pp. 172–203.
29. Fugazza, C.; Luraschi, G. Semantics-Aware Indexing of Geospatial Resources Based on Multilingual Thesauri: Methodology and Preliminary Results. *Int. J. Spat. Data Infrastruct. Res.* **2012**, *7*, 16–37.
30. Kuhn, W. Geospatial Semantics: Why, of What, and How? In *Journal on Data Semantics III*; Spaccapietra, S., Zimányi, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 1–24.
31. Kuhn, W. Modeling vs Encoding for the Semantic Web. *Semant. Web* **2010**, *1*, 11–15.
32. World Wide Web Consortium. SKOS Simple Knowledge Organization System Reference. 2009. Available online: <https://www.w3.org/TR/skos-reference/> (accessed on 5 November 2017).
33. Bechhofer, S.; van Harmelen, F.; Hendler, J.; Horrocks, I.; McGuinness, D.L.; Patel-Schneider, P.F.; Stein, L.A. OWL Web Ontology Language Reference. 2004. Available online: <https://www.w3.org/TR/owl-ref/> (accessed on 5 November 2017).
34. Lassila, O.; Swick, R.R.; World Wide Web Consortium. Resource Description Framework (RDF) Model and Syntax Specification. 1998. Available online: <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/> (accessed on 5 November 2017).
35. Berners-Lee, T. Linked Data—Design Issues. 2006. Available online: <http://www.w3.org/DesignIssues/LinkedData.html> (accessed on 5 November 2017).
36. Prud'hommeaux, E.; Seaborne, A. SPARQL Query Language for RDF. 2008. W3C. Available online: <http://www.w3.org/TR/rdf-sparql-query/> (accessed on 5 November 2017).
37. Chrisman, N. *Exploring Geographic Information Systems*; Wiley: Hoboken, NJ, USA, 1997.

38. Riedemann, C.; Pundt, H.; Harvey, F.; Kuhn, W.; Bishr, Y. Semantic interoperability: A central issue for sharing geographic information. *Ann. Reg. Sci.* **1999**, *33*, 213–232.
39. Martin, D.; Burstein, M.; Hobbs, J.; Lassila, O.; McDermott, D.; McIlraith, S.; Narayanan, S.; Paolucci, M.; Parsia, B.; Payne, T.; et al. OWL-S: Semantic Markup for Web Services. 2004. Available online: <http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/> (accessed on 5 November 2017).
40. Szekely, P.; Knoblock, C.A.; Gupta, S.; Taheriyan, M.; Wu, B. Exploiting Semantics of Web Services for Geospatial Data Fusion. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies (SSO '11), Chicago, IL, USA, 1 November 2011; pp. 32–39.
41. Dill, S.; Eiron, N.; Gibson, D.; Gruhl, D.; Guha, R. SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation. In Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 20–24 May 2003.
42. Mahmoudi, K.; Faiz, S. From Text to Semantic Geodata Enrichment. *Int. J. Agent Technol. Syst.* **2014**, *6*, 28–44.
43. Klien, E.; Lutz, M. The Role of Spatial Relations in Automating the Semantic Annotation of Geodata. In Proceedings of the International Conference on Spatial Information Theory (COSIT 2005), Ellicottville, NY, USA, 14–18 September 2005; pp. 133–148.
44. Klien, E. A Rule-Based Strategy for the Semantic Annotation of Geodata. *Trans. GIS* **2007**, *11*, 437–452.
45. Gomes de Andrade, F.; de Souza Baptista, C.; Henriques, H.B. Semantic Annotation of Geodata Based on Linked-open Data. In Proceedings of the 7th International Conference on Management of Computational and Collective Intelligence in Digital EcoSystems (MEDES '15), Caraguatatuba, Brazil, 25–29 October 2015; ACM: New York, NY, USA, 2015; pp. 9–16.
46. Vockner, B.; Mittlböck, M. Geo-Enrichment and Semantic Enhancement of Metadata Sets to Augment Discovery in Geoportals. *ISPRS Int. J. Geo-Inf.* **2014**, *3*, 345–367.
47. Nowak, J.; Nogueras-iso, J.; Peedell, S. Issues of multilinguality in creating a European SDI—The perspective for spatial data interoperability. In Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework, Alghero, Italy, 29 June–1 July 2005.
48. Li, W.; Goodchild, M.F.; Raskin, R. Towards geospatial semantic search: exploiting latent semantic relations in geospatial data. *Int. J. Digit. Earth* **2014**, *7*, 17–37.
49. Fox, P.; McGuinness, D.; Cinquini, L.; West, P.; Garcia, J.; Benedict, J.; Middleton, D. Ontology-supported Scientific Data Frameworks: The Virtual Solar-Terrestrial Observatory Experience. *Comput. Geosci.* **2009**, *35*, 724–738.
50. Li, W.; Yang, C.; Raskin, R. A semantic enhanced search for spatial Web portals. In *AAAI Spring Symposium—Technical Report*; AAAI: Palo Alto, CA, USA, 2008; Volume SS-08-05, pp. 47–50.
51. Akanbi, A.K.; Agunbiade, O.Y.; Kuti, S.; Dehinbo, O.J. A Semantic Enhanced Model for effective Spatial Information Retrieval. In Proceedings of the World Congress on Engineering and Computer Science (WCECS 2014), San Francisco, CA, USA, 22–24 October 2014.
52. Tagliolato, P.; Oggioni, A.; Fugazza, C.; Cianferoni, F.; De Fellici, S. Georiferimento di campioni museali nell'infrastruttura LifeWatch Italia: Le nuove prospettive dal web semantico. In Proceedings of the XXVI Congresso ANMS, Trieste, Italy, 16–18 November 2016.
53. Fugazza, C.; Pepe, M.; Oggioni, A.; Tagliolato, P.; Pavesi, F.; Carrara, P. Describing Geospatial Assets in the Web of Data: A Metadata Management Scenario. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 229.
54. XSL Working Group and the XML Linking Working Group. XML Path Language (XPath) Version 1.0. 1999. Available online: <https://www.w3.org/TR/xpath/> (accessed on 5 November 2017).
55. Open Geospatial Consortium Inc. (OGC). *Semantic Annotations in OGC Standards*; Discussion Paper OGC 08-167r1; Open Geospatial Consortium: Wayland, MA, USA, 2009.
56. Web Annotation Working Group. Web Annotation Vocabulary. Available online: <https://www.w3.org/TR/annotation-vocab/> (accessed on 5 November 2017).
57. Bigagli, L.; Mazzetti, P.; Nativi, S.; Boldrini, E.; Papeschi, F. GI-Cat: A Mediation Solution for Building a Clearinghouse Catalog Service. In Proceedings of the 2009 International Conference on Advanced Geographic Information Systems & Web Services (GEOWS 2009), Cancun, Mexico, 1–7 February 2009; Volume 00, pp. 68–74.
58. Bigagli, L.; Nativi, S.; Mazzetti, P.; Villoresi, G. GI-Cat: A Web Service for Dataset Cataloguing Based on ISO 19115. In Proceedings of the 15th International Workshop on Database and Expert Systems Applications (DEXA 2004), Zaragoza, Spain, 30 August–3 September 2004; pp. 846–850.

59. Linked Data Platform Working Group. Linked Data Platform. 2015. Available online: <https://www.w3.org/TR/ldp/> (accessed on 5 November 2017).
60. Web Annotation Working Group. Web Annotation Protocol. Available online: <https://www.w3.org/TR/annotation-protocol/> (accessed on 5 November 2017).



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).