Article

Raising Semantics-Awareness in Geospatial Metadata Management

Cristiano Fugazza [1,*] , **Monica Pepe** [1] , **Alessandro Oggioni** [1] and **Paolo Tagliolato** [1,2] and **Paola Carrara** [1]

[1] Institute for Electromagnetic Sensing of the Environment, National Research Council (IREA-CNR), v. Bassini 15, 20133 Milan, Italy; pepe.m@irea.cnr.it (M.P.); oggioni.a@irea.cnr.it (A.O.); tagliolato.p@irea.cnr.it (P.T.); carrara.p@irea.cnr.it (P.C.)

[2] LifeWatch Italy, Italy

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

Abstract: Geospatial metadata are often encoded in formats that either are not aimed at efficient retrieval of resources or are plainly outdated. Particularly, the quantum leap represented by the Semantic Web did not induce so far a consistent, interlinked baseline in the geospatial domain. Datasets, scientific literature related to them, and ultimately the researchers behind these products are only loosely connected; the corresponding metadata intelligible only to humans, duplicated in different systems, seldom consistently. We address these issues by relating metadata items to resources that represent keywords, institutes, researchers, toponyms, and virtually any RDF data structure made available over the Web via SPARQL endpoints. Essentially, our methodology fosters *delegated* metadata management as the entities referred to in metadata are independent, decentralized data structures with their own life cycle. Our example implementation of delegated metadata envisages: (i) editing via customizable web-based forms (including injection of semantic information); (ii) encoding of records in any XML metadata schema; and (iii) translation into RDF. Among the semantics-aware features that this practice enables, we present a worked-out example focusing on automatic update of metadata descriptions. Our approach, demonstrated in the context of INSPIRE metadata (the ISO 19115/19119 profile eliciting integration of European geospatial resources) is also applicable to a broad range of metadata standards, including non-geospatial ones.

Keywords: Spatial Data Infrastructures; geospatial metadata; RDF; semantics

1. Introduction

In the last decade, terms including Open Data [1], Linked Data [2], RDF [3], and SPARQL [4] have been more and more associated with open-ness and data interoperability. In fact, the emergence of Semantic Web technologies [5,6] established new practices for addressing two essential components of data management at large (i.e., data representation and access) and also provided means for (limited) processing of information by harnessing *ontologies* for formal description of schemata. To address management of geospatial resources in the framework proposed by the Semantic Web, metadata management cannot disregard the novel techniques that are elicited by the standards and practices defined in this context. Unfortunately, these technologies did not induce so far a consistent, interlinked baseline in the geospatial domain.

In this paper, we propose a methodology for *delegated*, evolving metadata that combines compliance with the normative standards and novel semantics-aware features: Native RDF representation allows for turning single-tenanted, monolithic metadata records into *living documents*. First, our methodology allows *normalizing* metadata descriptions by pinpointing the entities

(individuals, institutions, terms from controlled vocabularies, toponyms, etc.) that can be mapped to unique, externally managed resources identified with URIs [7] (or a specific fragment of these resources) and expressed as RDF. Instead, current metadata management practices express multiple references to each of these entities as independent metadata fragments that require manual update and are prone to inconsistencies and heterogeneities.

Semantic characterization of metadata is essential because, despite its structured encoding, resource descriptions often fall short of providing the intended interoperability because of the textual nature of property values. Recourse to code lists (i.e., sets of terms that represent the controlled vocabularies for specific metadata fields), as opposed to free text, clearly constitutes a sub-optimal solution. Moreover, not all property values can be constrained by code lists (e.g., the e-mail associated with a point of contact) and, nevertheless these can be related to unambiguously identified and authoritative data structures: As an example, a point of contact could be expressed by the URL of the ORCID profile (http://orcid.org/) of the researcher, whose information could then be retrieved, through content negotiation, as RDF.

Beside normalization, another important achievement of our research is *delegation*, that is, referring to external data structures for specification of a metadata property value: This is a direct by-product of the open-world vocation of URIs. We demonstrate the approach by means of the tools we developed in our projects, implementing ex-ante semantic lift of metadata. In this paper, among the advantages of semantics-aware metadata, we show that this methodology enables a paradigm shift in preserving metadata consistency. Specifically, we demonstrate that it is possible to automatically update individual metadata items, as well as more complex metadata structures, by associating the entities that are referred to in metadata with URIs and extracting individual property values at metadata-request time with SPARQL queries.

Further abstracting from schema-dependent metadata representation, our methodology also allows descriptions that are based on different schemas to refer to the same resources and extract the appropriate information subset: As an example, the fragments defining a point of contact in INSPIRE [8–10] and SensorML [11] metadata descriptions can be consistently drawn from the same authoritative source, even if they feature different components. Native XML metadata only allows for normalized representation on a per-schema basis by using XLink [12] techniques. Finally, notwithstanding the overall decentralized approach to metadata, our methodology allows retaining control on those information items that need to be controlled (Section 3.2 shows that it is possible to interleave proprietary RDF data structures in order to restrain the effects of delegation).

The remaining of the paper is organized as follows. Section 2 introduces the issues that stimulated this research and highlights the limitations of metadata representation techniques currently in use in the geospatial domain. Then, it presents relevant related work and a use case that typically leads to metadata inconsistency unless manual update is executed catalog-wide. Section 2.4 presents the application we used for injection of semantic information in metadata at editing-time and the workflow we implemented. Section 3 presents a general, qualitative evaluation of the possible usages of delegation in metadata management and then focuses on the use case that has been proposed. Specifically, Section 3.1 demonstrates that this practice can be employed to address automatic update of metadata items while Section 3.2 refines the approach, relaxing delegation for a specific metadata item while preserving its advantages. This is achieved by adding a further indirection step to augment flexibility in articulating organizational roles. Section 4 generalizes the approach to application domains different from the geospatial one, pinpoints the components that are necessary to implement delegation with a different application stack, and elaborates on the advantages of implementing the approach in a pure Linked Data scenario. Finally, Section 5 draws conclusions and outlines future work.

2. Methodology

This work can be broadly ascribed to the context of geospatial semantics. In particular, it relates to *encoding* of metadata in a semantics-aware fashion, as opposed to *modeling* of data and metadata

by using ontologies [13]. Consequently, important features that are in the scope of the latter, such as inference and subsumption, 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 [14] rather than a full-fledged OWL ontology [15]) is important because it determines 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).

Metadata are essential to effective provisioning of geospatial information through an SDI [16–18]. Consequently, much effort has been devoted in the last twenty years to the creation of metadata schemas that could encompass the variety of data categories and requirements by the stakeholders, from data providers to end users; a comprehensive but rather outdated view on these can be found in [19,20]. However, elaborating on the differences between the many metadata schemas that have been defined in the geospatial domain is out of the scope of this work, primarily because its aim is to abstract from the specific schema that is required by a given application scenario. The development of software tools for editing and maintenance of metadata progressed in parallel—a good survey of the available solutions is that provided by the US Federal Geographic Data Committee-Geospatial Metadata Tools (FGDC: https://www.fgdc.gov/metadata/geospatial-metadata-tools) but this activity is hampered by the intrinsic complexity of this task. In fact, it is difficult to conceive a one-size-fits-all solution in a landscape where requirements, established practices, and de-facto standards easily yield to custom profiles. We tackle schema heterogeneity issues by providing a schema-agnostic solution that can be tailored to the specific domain under consideration.

Assisted metadata creation is the approach we implemented to transparently associate URIs with metadata properties. As opposed to automatic metadata generation, such as the techniques proposed in [21–23], assisted metadata editing allows complementing the metadata properties that can be directly extracted from the dataset itself with the thematic and attribution characterizations that are very important for discovery and citation purposes. Instead, the aforementioned works rely on placement of datasets in specific sub-folders of a directory tree for thematic characterization of datasets (a practice we deem as error prone as manual specification of the corresponding metadata property values). In addition, points of contact are automatically derived from system-wide configuration of catalog software, a practice that cannot address attribution at the required granularity level (e.g., consider the management of shared SaaS catalogs).

In the development of our tools for assisting metadata creation, we investigated a broad range of schemas, from Unidata's NcML (the NetCDF Markup Language (NcML): http://www.unidata.ucar.edu/software/thredds/current/netcdf-java/ncml/ exploited within THREDDS servers for the annotation of NetCDF resources, to the OGC SWE family and in particular SensorML, the XML language aimed at the description of sensors. This choice of example schemas helps to mark the two ends of such spectrum: NcML, to enable some degree of interoperability among catalogs and resources, relies on the CF conventions [24] but, at the end of the day, metadata present code values as plain text; SensorML definition, instead, provides several XML attributes for complementing textual property values with unambiguous identifiers, such as URIs.

In this paper, we apply our methodology to the INSPIRE profile of the reference ISO standard as an in-between example of metadata format. On the one hand, it provides several code lists (defined in the schema itself or inherited from ISO 19115/19119) for property values. On the other hand, code values must be represented in metadata as plain text. It should be noted that recent updates to the technical guidelines for creation of INSPIRE metadata [25] allows for complementing any plain-text property value with URIs (included as *xlink:href* attributes). Nevertheless, adequate support for this functionality in editors is still missing.

To our knowledge, no prior works addressed conjunct production of heterogeneous geospatial metadata (to populate geoportals) and their RDF representation (aimed at extending metadata management capabilities). Although many works propose techniques for exposing geospatial metadata (complying with selected profiles) as RDF (e.g., [26]), we claim that only native RDF representation

can prevent the information loss that is typical in metadata crosswalks. This is the approach followed in [27], which also hints at the benefits of normalization that are inherent to this approach. However, the next section draws the state of the art with regard to metadata provisioning, the starting point of our implementation of the delegation paradigm. In doing this, we focus on independence of a specific schema and "pluggability" of external RDF data sources. The former is a requirement in many application scenarios. The latter is a prerequisite to relating information items to heterogeneous resources in the Web of Data.

2.1. Related Works: Schema Independence

From a theoretical viewpoint, the approach implemented in the CatMEDit editor (http://catm dedit.sourceforge.net/) and described in [28] is valuable because it addresses schema heterogeneity issues by harnessing the MDA techniques developed by Object Management Group-Model Driven Architecture (OMG: http://www.omg.org/mda/) In a nutshell, in this work, the authors abstract from the specific metadata schema by defining distinct *metamodels* that can then be translated into the corresponding editing interface, reflecting all the constraints defined by the former. In principle, this can address any possible custom profile derived from ISO 19115/19119 but requires background knowledge on the distinct phases that, in MDA, allow for translating the abstract model into implementation. Unfortunately, geospatial projects are increasingly requiring support to multiple metadata schemas for implementation, which calls for major modifications to the application. In addition, in the pragmatic framework of our research, we opted for a gentler learning curve, decoupling customization from implementation by requiring modification of a *template*, a schema definition, rather than of the application itself. Moreover, we wanted to develop a SaaS solution (although our editor can be used as a desktop application as well) while CatMEDit is solely available as a desktop application.

In line with this vocation, among the many products available in the state of the art for the provision of geospatial metadata (please refer to the aforementioned survey by FGDC for a list of these), the one provided by the GeoNetwork open source project software (http://geonetwork-ope nsource.org/), version 2.8 and later, is to our knowledge the only tool allowing for easily "pluggable" metadata schemas, but lacks support for external semantic sources. We can still consider the tools that, separately, address the most widely acknowledged metadata schemas for geographic and observational data, that is, ISO 19100 profiles and SensorML descriptions. The use case presented in this work is focusing on the first category of metadata schemas, which is also the one most widely supported by editing tools. The survey by FGDC makes it apparent that, whereas mature editors provide advanced features, existing ISO metadata editors only have partial support to customization of the governing metadata schema and no support at all for third-party data sources.

The same applies to the second category of metadata editing tools, that is, those devoted to SensorML. A similar review can be found on the OGC website at the page dedicated to SensorWeb Enablement Software (http://www.ogcnetwork.net/SWE_Software). A more up to date list of tools can be found in deliverable D3.3 of the ODIP 2 project (http://www.odip.eu/media/odip/org/doc uments/d3-3-progress-prototypes-1-submitted.pdf). The two editing tools referred to in the first review are the Pines SensorML Editor (http://lxspine.googlepages.com/pine'ssensormleditor), and the SensorML Process Editor (http://code.google.com/p/sensorml-data-processing/). Both are implementing an outdated version of SensorML and none implements the functionalities implied by our requirements. Considering the editing tools only included in the second list, such as the OpenSensorHub SensorML editor (https://github.com/opensensorhub/sensorml-editor) and the SensorNanny-DrawMyObservatory editor (https://github.com/ifremer/snanny-drawmyobservat ory), none of them provides such flexibility with respect to the metadata schema that is implemented. More importantly, plugging in data structures provided by third-party services is another functionality not implemented in existing editors.

2.2. Related Works: Normalization and Delegation

The resource-oriented vocation of ISO 19115/19119 metadata [29] and the graph-oriented capabilities of GML [30] suggest that normalization of metadata has been in the air long before this paper. Metadata schemas that are harnessing the W3C XLink technology, such as SWE-related standards and GML, or that have been extended to do so, can easily realize a similar normalization (and delegation) philosophy. In fact, metadata implementing these schemas could be stored in fully normalized fashion (i.e., substituting multiple references to the same metadata item, such as the aforementioned specification of a point of contact, with XLink references to a single metadata fragment) and use an XLink processor to generate the denormalized version of the record.

Unfortunately, this can only be done on a per-schema basis because, as soon as support for a different metadata schema is required (e.g., SensorML in conjunction with INSPIRE metadata), a second metadata fragment shall be mint and updated separately from the former. Our methodology realizes full normalization of metadata items by combining identification via URIs, RDF representation of information items, and a template language specifying how to retrieve individual property values for producing the intended denormalized output. To pinpoint the research problem we are addressing, we provide a use case exemplifying consistency issues that, in traditional metadata management, call for manual update of metadata records. We are going to see that, by adopting our metadata representation paradigm, it is possible to address these issues in an automated fashion.

As already mentioned, normalization of metadata records is among the achievements of [27]. Drawing inspiration from the digital library domain (a more mature context with respect to the geospatial one), the author articulates native RDF descriptions and provides multi-format translations that are uniquely identified as composite objects. However, this work focuses on schema selection and access protocols, while our work primarily focuses on referring to instances of RDF schemata and ontologies as metadata property values. Normalization is also the topic of [31] as a prerequisite to effective metadata maintenance. The paper proposed three approaches for achieving this but the first is strictly proprietary while the others are ex-post techniques aimed at system administrators as they involve the XSLT and Python scripting/programming languages. Moreover, in this work normalization does not imply (open-world) unique identification of the principals involved (researchers, institutes, and keywords) and this impedes delegation.

2.3. Use Case: John Doe Changes Employer

Since 2007, the implementation of the INSPIRE Directive indicated interoperability recipes to the geospatial community, basing them on international standards. Since the formulation of the Directive, however, the data management landscape has inevitably changed. The INSPIRE prototypical implementation with respect to metadata has been driven by the reference ISO 19100 standard series for geographic information [32–34], but the latter has become inadequate, mostly because of the increasing focus on data access and processing, as opposed to data representation (i.e., the search for mechanisms to appropriately encode data and metadata with respect to a specific domain).

It should be noted that this standard series also constituted the starting point for non-European standards, such as the NAP [35], the PCGIAP [36], and the ANZLIC [37] metadata profiles. Moreover, in the case of adoption or recommendation of different metadata standards, the ISO 19100 standard series is still acknowledged as the reference specification for Geographic Information: As an example, both the NASA ECHO Metadata Standard (https://earthdata.nasa.gov/standards/echo-metadata-standard) and the Directory Interchange Format of GCMD (http://gcmd.nasa.gov/add/standards/) are mapped to the core ISO elements. Finally, ISO metadata is the schema implemented in widely acknowledged geoportals, such as GeoServer (http://geoserver.org/) and GeoNode (http://geonode.org/). Hence, it makes sense questioning the aptness of these standards for tackling the aforementioned new challenges in geospatial metadata management.

To exemplify and test our approach in the context of this widely acknowledged standard, we consider the following use case: The institute "ACME Research" manages geospatial resources

through an ISO-compliant catalog service. Its researcher "John Doe" is the creator of "Dataset ABC". He is also the custodian of all datasets within the institute's SDI, hence he is referred to twice in the metadata record associated with "Dataset ABC". Figure 1 provides a visual representation of the XML nodes indicating "John Doe" as both the creator and the custodian of the resource in ISO metadata. Note that arcs abridge the traditional verbosity of nested XML structures in ISO 19139 encoding; in addition, Figure 1 highlights only the metadata items that are going to require update in the worked-out example presented in this paper.

"John Doe", at some point, changes employer and hence the metadata records of all datasets provided by the SDI of "ACME Research" are made inconsistent: Specifically, when referring to him as the creator of "Dataset ABC", his e-mail address and the name of the institute he works for have to be changed. Moreover, the company is likely to appoint a new custodian for the resources in the SDI, requiring update of all metadata records. We keep quotes when indicating the three entities involved in the use case (i.e., the individual, the institute, and the dataset) to stress that what one may intuitively consider complex data structures are actually mere character strings in (traditional) metadata records.

Such update must be manually carried out, a process that can be prone to errors or simply overlooked. Instead, the worked-out example in the following of this paper involves unambiguous identification of both the individual and his new employer: This allows up-to-date property values to be retrieved on demand as soon as the metadata record is requested.
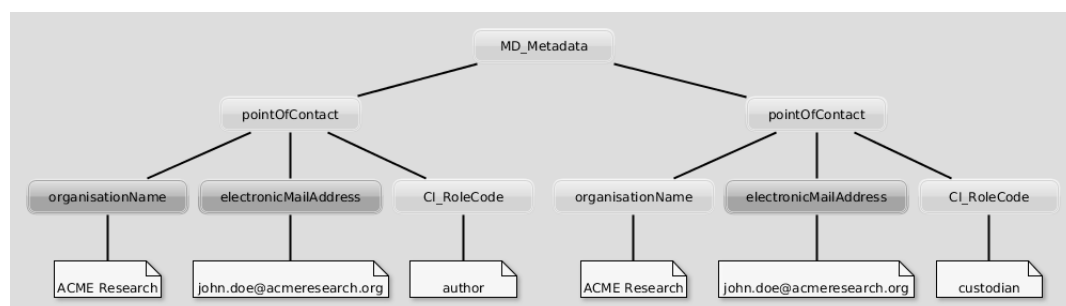

Figure 1. Nodes in ISO-compliant metadata indicating "John Doe" as both creator and custodian of resource "Dataset ABC'.

Rationale for Use Case Selection

The requirements that drove selection of an appropriate use case were the following:

1. The use case shall feature metadata items whose update can be triggered by an agent that is completely detached from the system hosting the metadata records themselves. Otherwise, one could argue that automatic update can be carried out by some procedure or daemon residing on this system.
2. The update shall affect multiple metadata records in order to show the breadth of automatic update.
3. Finally, the use case shall make it possible to exemplify the indirection mechanism presented in Section 3.2.

It is apparent that those metadata items that are overtly geospatial, such as spatial and temporal extents, violate all requirements. On the one hand, their update typically is due to addition of data instances in a series, an activity that can easily trigger update of the corresponding metadata. On the other hand, their update is likely to affect a single dataset. Moreover, they can not exemplify indirection. Another essential characterization of geospatial metadata is the thematic component expressed by keywords [38]. These are prone to change due to versioning of the corresponding thesauri or ontologies but whether (and how) they shall change in the metadata containing them is questionable. The remaining metadata item that is crucial for assessing aptness of a geospatial dataset, lineage information, is amenable to delegation but violates requirement 3.

2.4. The Proposed Approach

In this Section, we show how the tools we developed have been articulated in a workflow allowing metadata maintainers for the editing of descriptions that refer to external resources identified by URIs. Although different methodologies can produce similar results, such as the semantic lift of pre-existing metadata we investigated in [39], ex-ante characterization of metadata is more reliable with respect to the associated RDF resources. In addition, the second part of the workflow we describe in this Section is specifically aimed at addressing the inconsistency issues in the use case.

The core ingredient of our implementation of metadata delegation is EDI (http://edidemo.get-it.it) [40], a metadata editing tool for the production of metadata records in accordance to generic XML-based schemas by relying on a metalanguage: The EDI template language, a formalism defined by an appropriate XML Schema [41]. Specifically, EDI templates express custom metadata schemas that drive creation of web-based authoring interfaces to assist the user in providing the metadata. The client-side component of EDI is a JavaScript application that connects to SPARQL-compliant endpoints to ground autocompletion functionalities and context-based validation. Figure 2 shows the "Responsible party" section of the editing interface for INSPIRE metadata. It illustrates the three main features for assisted metadata editing that leverage external data sources: (i) autocompletion functionalities; (ii) conditional automatic compiling of fields; and (iii) dynamic populating of dropdown lists. These fields can be used to generate the metadata items in Figure 1 and are created on the basis of the template definitions in Listing 1.

Figure 2. Assisted editing of the "Responsible party" section of INSPIRE metadata.

Listing 1. Code fragment from the INSPIRE template.

```
1   <template>
2     <datasources>
3       <sparql xml:id="person">
4         <query><![CDATA[
5           SELECT ?contact ?label
6           WHERE {
7             ?contact  rdf:type      foaf:Person .
8             ?contact  vcard:email  ?label .
9             FILTER( REGEX( STR(?label), "$search_param", "i") ) }
10          ORDER BY ASC(?label)
11        ]]></query>
12        <url>http://url.to.endpoint/</url>
13      </sparql>
14    </datasources>
15    <element xml:id="resp">
16      <label xml:lang="en">Responsible party</label>
17      <produces>
18        <item hasIndex="1" datasource="person">
19          <label xml:lang="en">Email</label>
20          <hasPath>/.../gmd:electronicMailAddress/...</hasPath>
21          <rdfIn><![CDATA[
22            ?contact     vcard:email  ?1 .
23            ?contact     vcard:org    ?institute .
24            ?institute  foaf:name    ?2 .
25          ]]></rdfIn>
26        </item>
27        <item hasIndex="2" datasource="personDetails" field="name">
28          <label xml:lang="en">Institute</label>
29          <hasPath>/.../gmd:organisationName/...</hasPath>
30        </item>
31        <item hasIndex="3" datasource="roleCodes">
32          <label xml:lang="en">Role</label>
33          <hasPath>/.../gmd:CI_RoleCode/...</hasPath>
34          <rdfIn>...</rdfIn>
35        </item>
36      </produces>
37      <rdfOut><![CDATA[
38        INSERT {
39          <$id_1_uri> dcat:contactPoint [
40            vcard:hasUID    <$resp_1_uri> ;
41            vcard:hasRole   <$resp_3_uri> ] . }
42        WHERE {}
43      ]]></rdfOut>
44      <rdfIn><![CDATA[
45        SELECT ?contact ?1 ?institute ?2
46        WHERE {
47          <$id_1_uri>  dcat:contactPoint  ?struct .
48          ?struct       vcard:hasUID       ?contact . }
49      ]]></rdfIn>
50    </element>
51  </template>
```

The best deployment option for EDI is as part of the workflow set by the catalog software of choice for the provision of datasets. In fact, this strategy allows system administrators to narrow the number of property values that shall be provided by metadata maintainers. On the one hand, naming conventions used by the catalog software can be matched in template structures to keep some fields, such as resource identifiers, transparent to the end user. On the other hand, a number of fields (e.g., the spatial extent of the dataset) can be computed at resource upload-time and passed on to the editing interface as request parameters. As an example, integration of EDI with GeoNode allowed us to dramatically narrow the number of required fields for the provision of RNDT metadata [42] (the Italian profile of INSPIRE metadata): Out of 34 mandatory fields, the number of those that the metadata maintainer has to provide from scratch was reduced to six.

A server-side component, written in Java, executes the actual translation of the user input in both the XML and RDF representations that allow us to demonstrate metadata delegation in this paper. This tool (please refer [40,43–45] for further information) allows for a threefold exploitation of RDF data structures; namely:

- It is possible to plug in generic RDF data sources available as SPARQL endpoints. Based on query results, property values of metadata items can be filled in.

- A semantics-aware version of the metadata record can be created and encoded as RDF, beside the XML profile the editing tool is configured to produce.
- The metadata in the original XML profile can be reconstructed on demand (refer to the use case in the previous Section), thus relying on up-to-date information for reducing as possible inconsistency issues.

A mantra in contributing to the Semantic Web is reuse of existing schemata. Hence, for the translation into RDF of ISO-based geospatial metadata, our primary reference was the OWL representation of the ISO/TC 211 UML Model for geographic information developed by CSIRO [46]. The W3C vocabulary DCAT [47], the data schema underlying the existing Open Data portals based on the open source software CKAN (http://ckan.org/), constituted another source of direction. Moreover, the DCAT application profile for data portals in Europe [48] has been recently complemented by an extension for representing INSPIRE metadata as RDF [49]. Having in mind to ease integration with CKAN-based portals, we adopted the DCAT schema as the baseline but, when choosing between the aforementioned contributions to the state of the art to express geospatial-related metadata, we opted for [46]. In fact, the GeoDCAT extension [49] is primarily aimed at integration with CKAN while we need broader support for URI-based metadata property values.

2.5. Semantic Lift of Resource Metadata

Semantic lift of metadata essentially amounts to relating data items in metadata records to entities in the Web of Data through their identifiers, their URIs. This is done at metadata creation-time via EDI: On the one hand, the tool draws information from SPARQL endpoints for assisting editing of metadata and, on the other hand, the tool stores the URIs of the entities that are selected by the metadata maintainer for future use.

Of course, this does not imply that all fields in a metadata description shall be decentralized in this fashion. In fact, we acknowledge that a number of these should change only as a consequence of voluntary update by the metadata maintainer. Nevertheless, there are metadata fields that are plainly useless or even misleading if not consistent (e.g., the details regarding a point of contact) and could therefore take great advantage of the delegation mechanism presented in this paper. The proposed solution does not envisage versioning of the metadata descriptions, of the data structures referred to by the former, as well as of the individual data items that are included in metadata. However, it should be noted that unique identification of the entities under consideration is a prerequisite to all these functionalities that can then be built on top of the RDF data structures.

Similarly, if concerns on the availability of the data structures that are referred to by metadata descriptions are raised, one could easily apply a caching layer so that, whenever a given value cannot be retrieved from the data source, the value can be taken from the static metadata description that have been previously generated and stored. In fact, if a given metadata item is related to a URI, it means that the identified data structure has been found at least once, at metadata creation-time, and then the corresponding static XML output can be stored and subsequently looked up if something goes awry.

In the following, we use the application described above to provide a prototypical implementation of the delegation principle presented in this paper. Figure 3a depicts the workflow that allow metadata maintainers to provide semantics-aware metadata. Albeit developed in the context of one of our projects, the methodology can be customized to any XML-based profile of geospatial metadata. To do so, a necessary preliminary activity is constituted by tailoring of the EDI template reflecting the metadata schema that shall be produced by the tool (Figure 3a, 1) Specifically, the system administrator (the person customizing EDI for the specific application scenario) can create the template from scratch, supported by the XML Schema provided in the software package. Alternatively, the administrator can customize one of the example templates (https://github.com/SP7-Ritmare/EDI-NG_templates) which implement a number of different INSPIRE and SensorML metadata profiles.

The template can now be fed to EDI to create the web form that metadata maintainers can use to edit metadata records (Figure 3a, 2).The form is going to reflect the constraints encoded in the

template (e.g., mandatory fields, multiplicities, data types, etc.) and connect to the remote data sources (Figure 3a, 3). Once the data are posted to the server-side component of EDI, the XML metadata record is generated alongside the corresponding RDF representation (Figure 3a, 4). The XML files can then be passed on to geospatial catalogs that can understand the specific formalism (in our scenario, these were either CSW catalogs or SOS servers) for traditional discovery: In this case, the advanced features described in this work can not be made available.
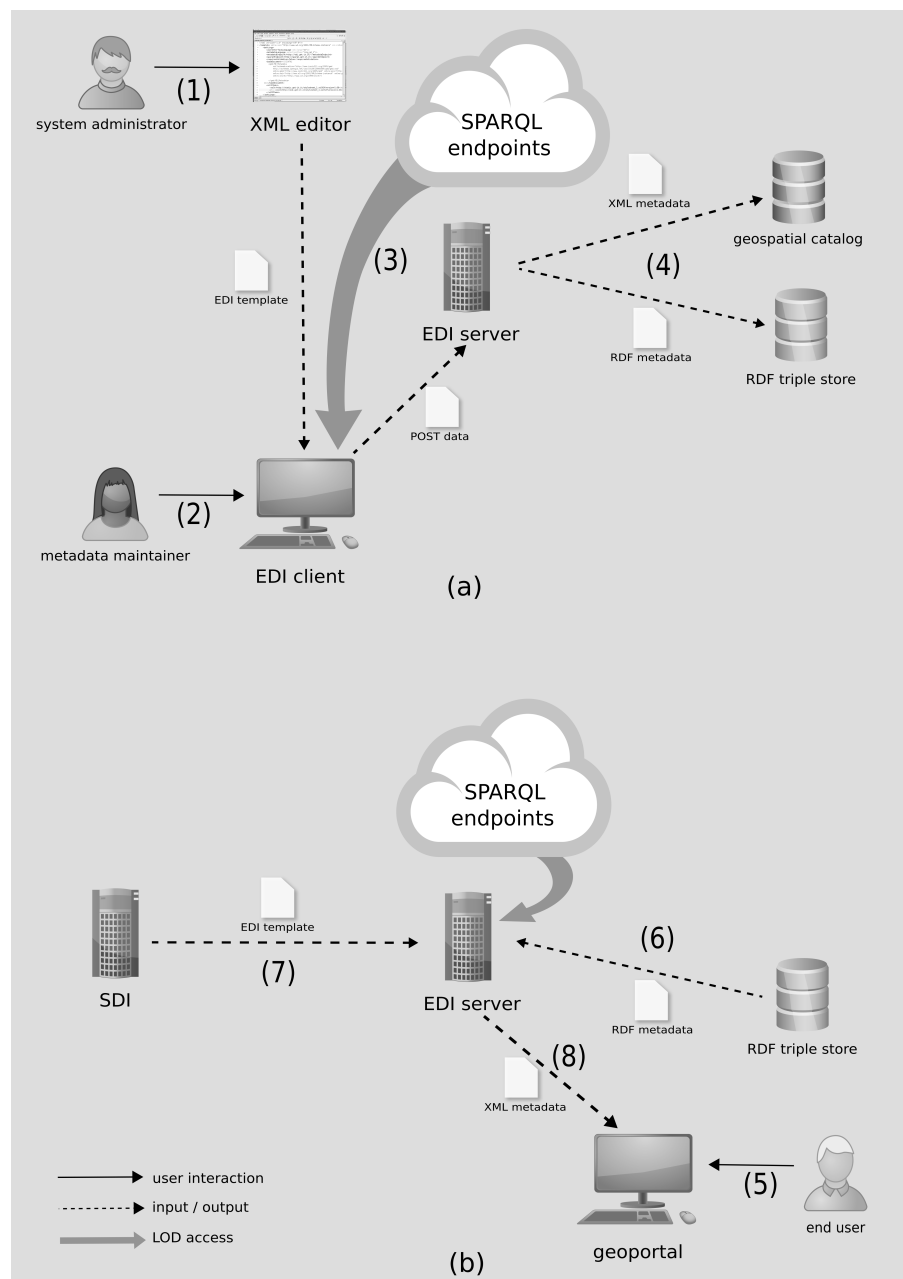

Figure 3. Workflow addressing: (**a**) production of the XML and RDF representations of the metadata record; and (**b**) on-demand generation of the XML representation.

In the fragment of the template for INSPIRE-compliant metadata (Listing 1), we highlight the logics underlying the three usages enumerated in the beginning of this Section. Note that the code is omitting a number of elements and attributes that, albeit necessary to enactment of the template, may be misleading to the reader. Specifically, Lines 3–13 define a datasource named "person" by specifying the SPARQL query to be issued to the endpoint specified on Line 12. The datasource is later

referred to on Line 18, in the definition of the metadata field corresponding to the e-mail address of the point of contact, thus enabling autocompletion of the field on the basis of query results. Specifically, parameter "$search_param" will be replaced run time by the text pattern entered by the metadata maintainer. In addition, the metadata item on Lines 27–30 will be autocompleted on the basis of the specific point of contact that is selected (see Figure 2, ii). Moreover, the field on Lines 31–35 can be selected from a drop-down list of roles that is dynamically created according to datasource "roleCodes" (as in Figure 2, iii). The mechanics of completion of the last two fields is omitted in Listing 1 for brevity. Further assisted editing functionalities include autocompletion on the basis of HTTP parameters in the call to the editing interface, on-demand generation of field values (e.g., to generate unique identifiers), and combination of values from other metadata fields.

The server-side component of the tool takes care of generating the prescribed XML output on the basis of XPath expressions (such as those on Lines 20, 29, and 33). Additionally, the *rdfOut* element on Lines 37–43 specify the RDF triples that shall be produced and inserted in the triple store used for metadata storage. In the example, Lines 37-43 contain the *blueprint* of the RDF data structures expressing the point of contact in Figure 1. Specifically, parameter "$resp_1_uri" retrieves the first item in element "resp" (containing the e-mail address of the point of contact) and looks up the corresponding URI (the unique identifier of the selected individual). The e-mail of the point of contact, as well as the name of the institute he works for, are not comprised in the triples defined by the *rdfOut* element: These values will be retrieved on demand, when the metadata of the resource is looked up. At mere syntactic level, this is the major difference between our approach and any of the possible one-to-one translations of INSPIRE metadata into RDF, such as those in [46,49]. In the next Sections, we provide a qualitative analysis of the novel capabilities that are elicited by our metadata representation strategy.

3. Results

Unique identification of the entities related to metadata property values improves overall interoperability of SDIs: Beside easing processing by automated agents, this practice helps addressing discovery issues related to multilingualism and semantic heterogeneity. It should be noted that association of URIs with metadata items also constitute a pivotal prerequisite to further functionalities that are elicited by Semantic Web technologies, such as query expansion and inference. These are not considered in this work as they are dependent on the specific discovery facilities that are implemented on top of metadata.

The potential of multilingual queries combined with query expansion is exemplified in [50]: In that paper, a search pattern expressed in Arabic is first turned into a language-neutral URI by looking up a geospatial thesaurus. Then, the inner structure of the latter is exploited in order to enrich the search parameters with additional terms. Multiple language-dependent queries can now be formulated, one for each language supported by the thesaurus, and they are likely to return more results than the original one. In the scenario proposed in that work, this technique can only be applied to the search pattern entered by the user but it is apparent that native URI-based metadata provide the same benefits without requiring back translation into language-dependent terms (for traditional discovery).

Normalization of metadata is a direct by-product of unique identification of property values. In and of itself, this feature implies the improvements with respect to metadata maintenance described in [27,31]. However, we remove the issues due to proprietary software and technological barriers by providing, respectively, a FOSS solution and a user-friendly template language. Moreover, relying on SPARQL endpoints for the provision of the data structures that are referred to in metadata allows decoupling management of the latter, implemented by the catalog software of choice, and management of the RDF data structures, which can take advantage of the broader set of tools made available in the Semantic Web context.

The following focuses on the use case presented in Section 2.3, addressed by the workflow in Figure 3b and assuring consistency of metadata records by leveraging on the delegated metadata created in the first part of the workflow (Figure 3a). Specifically, once the end user has discovered

the dataset of interest and requests the corresponding XML metadata (Figure 3b, 5), the EDI server retrieves the RDF metadata (Figure 3b, 6) and looks up the template (Figure 3b, 7) to access the correct data structures by the SPARQL endpoints that are defined. Then, the up-to-date XML representation of the former can be generated (Figure 3b, 8).

3.1. Basic Metadata Delegation

Each metadata element definition in the template can specify a number of *rdfIn* nodes that specify the SPARQL queries that return the values for the distinct metadata fields. As an example, Lines 44–49 in Listing 1 contain the query fragment retrieving the metadata items that are stored in the local triple store. Instead, individual *item* nodes define the query fragments that shall be executed against the remote endpoint that is defined for the specific *item*. Specifically, Lines 21–25 define the triples that shall be matched against the endpoint specified on Line 12. Note that, in Listing 1, we omitted namespace declarations for the sake of conciseness; in addition, *rdfIn* elements only contain the triple patterns that are necessary to address the use case in Section 2. Listing 2 presents the query compiled from the information in Listing 1. This straightforward query is directed to a data source which is expected to model information using the *vcard* schema; it could however be formulated in a more general way, by allowing for alternative values, multiple schemata, etc.

With respect to the use case of John Doe changing institute (Section 2.3), our representation of metadata as RDF exploits the information generated by the *rdfOut* element on Lines 37–43 of Listing 1. Suppose that John Doe updated his FOAF profile in order to feature his new e-mail address and point to the FOAF profile of the new employer. Now, the metadata for "Dataset ABC" can be composed on the fly: The generated XML tree is analogous to that in Figure 1, except that the two inconsistent leaf nodes in the leftmost sub-tree are automatically updated to the new values. One may argue that, to enable update of the metadata associated with "Dataset ABC", John Doe is required to update his FOAF profile, a modification that is clearly not automatic. However, a closer look will show that this modification is completely independent of the workflow for managing metadata that is implemented by ACME Research (actually, the research institute may not be aware of any change in John Doe's FOAF profile).

Listing 2. SPARQL query compiled from template definitions.

```
1   SELECT ?contact ?1 ?institute ?2
2   WHERE {
3     <http://.../metadata/Dataset_ABC>      dcat:contactPoint      ?struct .
4     ?struct                                vcard:hasUID           ?contact .
5     OPTIONAL {
6       SERVICE <http://url.to.endpoint/> {
7         ?contact      vcard:email   ?1 .
8         ?contact      vcard:org     ?institute .
9         ?institute    foaf:name     ?2 .
10      }
11    }
12  }
```

3.2. Adding Indirection to Delegation

We have seen thus far that, by indicating the entities referred to in metadata via their URIs, it is possible to achieve on-demand generation of ISO-compliant metadata records that have higher quality assurance with regard to consistency of metadata fields. In fact, every time the description of resource "Dataset ABC" is requested, the e-mail address of the creator of the dataset and the name of the institute he works for are updated to their current values without requiring manual update of the metadata record every time these values change. Unfortunately, this functionality is, in and of itself, of little use for addressing the second inconsistency issue described in Section 2.3. In fact, unless all the metadata records in the SDI of ACME Research are updated, they are going to point to John Doe, the old custodian, and provide incorrect information (albeit up-to-date). This happens because the mechanics described so far work well for individual property values of the entities that

are referred to but fall short of extending their behavior to whole entities (e.g., the specification of the new custodian for ACME Research). Moreover, since basic delegation takes control over metadata property values out of the party maintaining metadata, there is a need for upholding this control on those information items that require this (e.g., custodian information in the use case we presented). A straightforward means to achieve this is disallowing delegation for these values, that is, reverting to good old plain-text metadata (an option our methodology does not rule out) but this would not retain the advantages described in the previous Section.

Instead, it is possible to achieve automatic update of custodian information in all metadata records by using an in-between data structure whose purpose is to act as a signpost for the person fulfilling this role. This can be achieved by minting a URI representing the custodian signpost (that, in the example, is set to "http://.../custodian"). This URI can then be associated with the real individual fulfilling this role: As an example, Listing 3 contains the SPARQL query appointing John Doe as custodian of the resources in the SDI of ACME Research. Note that the specific property expressing this relation between the abstract signpost and the actual individual (in the example, *rdfs:seeAlso*) can be whatever because it is just a convention between the local triple store and the template. Specifically, Lines 6 and 7 in Listing 3 retrieve the triple that specifies the current custodian, whoever she is, while Lines 2 and 3 appoint John Doe to this role. The second step is constituted by creating a datasource in the template that is capable of taking into account this new information, such as the one shown in Listing 4. The query is very similar to that of datasource *person* in Listing 1, except that the query is run against the local triple store (where the triple inserted by the query in Listing 3 is residing) and only then the remote endpoint is interrogated.

Listing 3. SPARQL query creating the triple enabling indirection

```
1   INSERT {
2     <http://.../custodian>  rdfs:seeAlso  <http://.../JohnDoe> .
3   }
4   WHERE {
5     <http://.../custodian>  rdfs:seeAlso  ?individual .
6   }
```

Listing 4. Datasource definition implementing indirection.

```
1   <sparql xml:id="custodian">
2     <query><![CDATA[
3       SELECT ?email
4       WHERE {
5         <http://.../custodian>    dfs:seeAlso    ?contact .
6         SERVICE <http://url.to.endpoint/> {
7           ?contact    rdf:type    foaf:Person .
8           ?contact    vcard:email ?email .
9         }
10      }
11    ]]></query>
12  </sparql>
```

Finally, the EDI template shall contain the *element* definition that is going to access the triple in Listing 3 according to the data source definition in Listing 4, as shown in Listing 5. Specifically, the first *item* definition refers to datasource "custodian" and variable "email" in Listing 4 to retrieve the e-mail of the individual acting as the custodian for the SDI. The second *item* (corresponding to the name of the research institute) is known in advance. Moreover, the content of the *rdfOut* tag and of the outermost *rdfIn* tag shall be modified accordingly: The former features the same triple patterns as in Listing 1 except that the property values are statically set; the second is modified to take into account the indirection step.

Back to the use case presented in Section 2; when John Doe changes workplace, it is sufficient to run the query in Listing 3, after substituting his URI with that of the new custodian. This is going to automatically update all custodian information in the metadata records hosted by the SDI. In this

second flavor of metadata delegation we present, one may argue that manual update is required to assign the role of custodian to the successor of John Doe, but a closer look is going to put this objection in perspective. Firstly, this manual update is actually pursued by ACME Research in order to keep control on who is the custodian for the whole SDI, a modification that is less prone to be overlooked. Secondly, this individual update is going to produce catalog-wide update of metadata that does not require modification of all XML metadata records. Finally, this update may easily be the consequence of a company workflow different from that of metadata management, such as the update of registry information describing the employees of ACME Research.

Listing 5. Template definitions implementing indirection.

```
1     <produces>
2       <item xml:id="cust_1" datasource="custodian"  field="email">
3         <label xml:lang="en">Email</label>
4         <hasPath>/.../gmd:electronicMailAddress/...</hasPath>
5         <rdfIn><![CDATA[
6           ?contact  vcard:email  ?1 .
7         ]]></rdfIn>
8       </item>
9       <item hasIndex="2" xml:id="cust_2">
10        <label xml:lang="en">Institute</label>
11        <hasValue>ACME Research</hasValue>
12        <hasPath>/.../gmd:organisationName/...</hasPath>
13      </item>
14    </produces>
15    <rdfOut><![CDATA[
16      INSERT {
17        <$id_1_uri> dcat:contactPoint [
18          vcard:hasUID   <http://.../signpost> ;
19          vcard:hasRole  <http://.../ResponsiblePartyRole/custodian> ] .
20      } WHERE {}
21    ]]></rdfOut>
22    <rdfIn><![CDATA[
23      SELECT ?contact ?1 ?institute ?2
24      WHERE {
25        <http://.../signpost>  rdfs:seeAlso  ?contact . }
26    ]]></rdfIn>
```

4. Discussion

The example provided in this paper can be replicated in different application scenarios as the metadata profiles that are made available by our tool are of general interest to the INSPIRE and SWE communities [51]. Moreover, it should also be noted that the issues our research aims at solving affect a broader range of metadata formats, thus hampering efficient retrieval of resources in a number of different domains. The schema-agnostic approach that has been exemplified with a specific use case in the geospatial domain can then be applied to heterogeneous domains by tailoring the metadata template to the latter and by selecting the appropriate semantic sources.

Besides improved discovery functionalities (such as those described in [52,53]), our methodology aims at fostering distributed metadata management. Specifically, adoption of RDF as the native metadata storage format and extensive usage of URIs as property values enable management of geospatial metadata in a decentralized, multi-tenanted fashion. In fact, since the Semantic Web is a distributed corpus of information whose data resources can often be directly accessed, via the associated URIs, as Linked Data, it makes little sense sticking to the notion of metadata as data structures that are single-handedly maintained by the institution providing a given resource. Instead, a distributed approach could induce a *metadata fabric* that is more resilient to change. In addition, referring to semantics-aware entities in metadata allows making the latter more usable through formal specification of information items via RDF schemas and ontologies. The big picture our research is contributing to, on which we elaborate in the following paragraph, is fully hypermedia-driven metadata.

Throughout this paper, we draw the RDF descriptions of information items from SPARQL endpoints, the rationale for this being expressiveness of the query language and the need for readily usable facilities for assisting provision of metadata. However, it is apparent that the delegation

paradigm can be employed to its full potential only in a pure Linked Data ecosystem. Referring to Listing 2, the query is assumed to return bindings for variables *?1* and *?2*, corresponding to the *hasIndex* attributes of the related *item*s in Listing 1. Should any of these values be missing (e.g., because of removal of the record associated with John Doe from the remote endpoint), the result set still contains the necessary information, the URI bound to variable *?contact*, to try a Linked Data lookup and retrieve the value of the missing properties. Similarly, John Doe's FOAF description is supposed to define the URI of the institute he works for, expressed in the example as a *vcard:org* property definition. Then, the approach described so far can be reiterated and the name of the institute can be retrieved. If a triple store is used to cache the Linked Data chunks corresponding to the FOAF descriptions of John Doe and his new employer, the remote endpoint in the SERVICE clause of Listing 2 (Line 7) can be replaced with the endpoint of this caching layer so that the query can return the intended property values.

The delegation principle can also be implemented in a metadata management framework and with tools that are different from those employed in the worked-out example we presented in this paper. In this case, it should be noted that relating metadata property values to URIs is not sufficient to pinpointing the required property values. The missing component is specification of the search pattern that in our example implementation is constituted by the SPARQL queries defined in the template. From-scratch implementation of metadata delegation may consider *property paths* as a means to specify the data fragments that shall be retrieved but, because of the low expressiveness of property paths in the SPARQL 1.1 specification, recourse to extensions, such as those described in [54,55], may be necessary.

Finally, referring to the use case that was presented, beside keeping the metadata record consistent (allowing researchers interested in the dataset to contact its creator), delegation is also functional to maintaining authorship on the dataset by John Doe. This cannot be enforced in state-of-the-art metadata management and nevertheless it is important to uphold this principle in the framework of the author's digital identity. In fact, researchers are increasingly appointing third parties for the purpose of shaping their identities on the Web (e.g., through ORCIDs) and then leveraging the data structures created by the latter is just one step farther. Essentially, researchers should start considering management of their digital identities as a task that is complementary to ordinary metadata provision. This practice allows the former to maintain full control over the products they are providing across heterogeneous geoportals.

5. Conclusions and Future Work

This paper addresses piecewise articulation of metadata descriptions based on heterogeneous resources in the Semantic Web. We demonstrate the advantages of delegated management of metadata with an example implementation set in the geospatial domain. In the worked-out example, besides traditional metadata representation that is processable by traditional geospatial catalogs, we generate a semantics-aware RDF representation to be used at a later stage for providing increased features. Among the many functionalities that this characterization can ground, we propose a use case addressing automatic update of metadata properties in either individual records and catalog-wide. However, it should be noted that adopting RDF as the metalanguage for resource description contributes to moving close to the long-term intent expressed in the abstract—that is, creating a *trait d'union* among datasets, the scientific literature related to them, and the researchers behind these products. State-of-the-art infrastructures rely on centralized metadata management and compliance with outdated representation formats for their implementation. Instead, fluidity of research environments calls for managing metadata at a finer granularity.

The software that we used for demonstrating the benefits of metadata delegation (i.e., the EDI framework) is made available as FOSS on GitHub, at the SP7 Interoperable Data Infrastructure for RITMARE repository (https://github.com/SP7-Ritmare). The advantages of recourse to semantic data sources for assisted editing of metadata is attested by adoption of the tool by a number of projects, namely the FP7 projects ERMES (http://www.ermes-fp7space.eu/it/homepage/), 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 Projects RITMARE (http://www.ritmare.it/) and NextData (http://www.nextdataproject.it/). The metadata descriptions produced in these projects constitute the metadata-base we are using to assess applicability of the practice described in this paper also to other use cases, such as multilingual discovery and query expansion.

Our current research is also focusing on extending the approach to encompass pre-existing metadata descriptions via ex-post processing [39] and on sharing of the semantic information that is produced for reuse by third parties, such as the principals that created the resource descriptions in the first place. The semantic lift task featured in that work repurposes the same template structures described in this paper in a "reverse engineering" fashion and is complementary to the approach presented in this work. Specifically, pre-existing metadata that have been processed that way can then be re-edited by using the same application we used to demonstrate delegation. Instead, sharing semantic information relates to provision of the semantics-aware portion of a metadata description according to the WADM format proposed by W3C [56]. In fact, while exhaustive translation of metadata into RDF is primarily meant for storage and serialization of resource descriptions, pinpointing the semantic information that is produced is crucial in the interplay with traditional discovery, as widely implemented in state-of-the-art geoportals.

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
CF	Climate and Forecast
CSIRO	Commonwealth Scientific and Industrial Research Organization
CSW	Catalogue Service for the Web
DCAT	Data Catalog Vocabulary
ECHO	Earth Observing System Clearing House
FOAF	Friend Of A Friend
FGDC	Federal Geographic Data Committee
FOSS	Free and Open Source Software
GCMD	Global Change Master Directory
HTTP	HyperText Transfer Protocol
INSPIRE	INfrastructure for Spatial InfoRmation in Europe
ISO	International Organization for Standardization
MDA	Model Driven Architecture
NAP-Metadata	North American Profile of ISO 19115:2003—Geographic Information—Metadata
NcML	The NetCDF Markup Language
OGC	Open Geospatial Consortium
OMG	Object Management Group
OWL	Web Ontology Language
PCGIAP	Permanent Committee on GIS Infrastructure for Asia and the Pacific
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

SaaS	Software-as-a-Service
SDI	Spatial Data Infrastructures
SensorML	Sensor Markup Language
SKOS	Simple Knowledge Organization System
SOS	Sensor Observation Service
SPARQL	SPARQL Protocol and RDF Query Language
SWE	Sensor Web Enablement
UML	Unified Modeling Language
URI	Uniform Resource Identifier
XLink	XML Linking Language
XML	eXtensible Markup Language
XPath	XML Path Language
WADM	Web Annotation Data Model
W3C	World Wide Web Consortium

References

1. Dawes, S.S. A Realistic Look at Open Data. Using Open Data: Policy Modelling, Citizen Empowerment, Data Journalism. W3C, 2012. Available online: http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf (accessed on 20 August 2018).
2. Heath, T.; Bizer, C. *Linked Data: Evolving the Web into a Global Data Space*, 1st ed.; Morgan & Claypool: San Rafael, CA, USA, 2011.
3. 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 20 August 2018).
4. Prud'hommeaux, E.; Seaborne, A. SPARQL Query Language for RDF. W3C Recommendation, 2008. Available online: http://www.w3.org/TR/rdf-sparql-query/ (accessed on 20 August 2018).
5. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]
6. Hitzler, P.; Krtzsch, M.; Rudolph, S. *Foundations of Semantic Web Technologies*, 1st ed.; Chapman & Hall/CRC: London, UK, 2009.
7. Berners-Lee, T.; Fielding, R.; Masinter, L. Uniform Resource Identifiers (URI): Generic Syntax, 1998. Available online: https://tools.ietf.org/html/rfc3986 (accessed on 20 August 2018).
8. 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 20 August 2018).
9. 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 20 August 2018).
10. 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. Available online: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32008R1205R%2802%29 (accessed on 20 August 2018).
11. Botts, M.; Robin, A. OGC® Sensorml: Model and Xml Encoding Standard. Open Geospatial Consortium. Available online: http://www.opengeospatial.org/standards/sensorml (accessed on 20 August 2018).
12. DeRose, S.J.; Maler, E.; Orchard, D.; Walsh, N. XML Linking Language (XLink) Version 1.1. Technical Report, 2010. Available online: http://www.w3.org/TR/2010/REC-xlink11-20100506/ (accessed on 20 August 2018).
13. Kuhn, W. Modeling vs. Encoding for the Semantic Web. *Semant. Web* **2010**, *1*, 11–15.
14. World Wide Web Consortium. SKOS Simple Knowledge Organization System Reference, 2009. Available online: https://www.w3.org/TR/skos-reference/ (accessed on 20 August 2018).

15. Bechhofer, S.; van Harmelen, F.; Hendler, J.; Horrocks, I.; McGuinness, D.L.; Patel-Schneider, P.F.; Stein, L.A. OWL Web Ontology Language Reference. Technical Report, W3C, 2004. Available online: http://www.w3 .org/TR/owl-ref/ (accessed on 20 August 2018).

16. Kim, T.J. Metadata for geo-spatial data sharing: A comparative analysis. *Ann. Reg. Sci.* **1999**, *33*, 171–181. [CrossRef]

17. Tsou, M.H. An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on the Internet. In *Geographic Information Science*; Egenhofer, M.J., Mark, D.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 313–332.

18. Limbach, T.; Krawczyk, A.; Surowiec, G. Metadata Lifecycle Management with GIS Context. In Proceedings of the 10th EC GI & GIS Workshop, ESDI State of the Art, Warsaw, Poland, 23–25 June 2004.

19. Moellering, H.; Aalders, H.J.; Crane, A. *World Spatial Metadata Standards: Scientific and Technical Descriptions, and Full Descriptions with Crosstable*; International Cartographic Association Series; Elsevier: New York City, NY, USA, 2005.

20. Nogueras-Iso, J.; Muro-Medrano, P.; Zarazaga-Soria, F. *Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval*; Springer: Berlin, Germany, 2005.

21. Batcheller, J.K. Automating Geospatial Metadata generation-An Integrated Data Management and Documentation Approach. *Comput. Geosci.* **2008**, *34*, 387–398. [CrossRef]

22. Batcheller, J.K.; Gittings, B.M.; Dunfey, R.I. A Method for Automating Geospatial Dataset Metadata. *Future Internet* **2009**, *1*, 28–46. [CrossRef]

23. Olfat, H.; Rajabifard, A.; Kalantari, M. Automatic Spatial Metadata Update: A New Approach. In Proceedings of the XXIV FIG International Congress, Sydney, Australia, 11–15 April 2010.

24. Eaton, B.; Gregory, J.; Drach, B.; Taylor, K.; Hankin, S.; Caron, J.; Signell, R.; Bentley, P.; Rappa, G.; Höck, H.; et al. NetCDF Climate and Forecast (CF) Metadata Conventions, 2011. Available online: http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.pdf (accessed on 20 August 2018).

25. European Commission. Technical Guidance for the Implementation of INSPIRE Dataset And Service Metadata Based on ISO/TS 19139:2007, 2016. Available online: http://inspire.ec.europa.eu/id/document/t g/metadata-iso19139 (accessed on 20 August 2018).

26. Reid, J.; Waite, W.; Butchart, B. Going for GOLD—Adventures in Open Linked Metadata. In Proceedings of the AGI GeoCommunity, Nottingham, UK, 20–22 September 2011. Available online: https://pdfs.semanticscholar.org/74cd/e4de72801786e04a4a924a450ee8371f460e.pdf?_ga=2.533 38344.1342966457.1534749721-1794176989.1522156097 (accessed on 20 August 2018).

27. Westbrooks, E. Distributing and Synchronizing Heterogeneous Metadata for the Management of Geospatial Information Repositories. In Proceedings of the International Conference on Dublin Core and Metadata Applications, Seattle, WA, USA, 28 September–2 October 2003; pp. 67–77. Available online: http://dcpapers .dublincore.org/pubs/article/view/735 (accessed on 20 August 2018).

28. Nogueras-Iso, J.; Latre, M.A.; BéJar, R.; Muro-Medrano, P.R.; Zarazaga-Soria, F.J. A Model Driven Approach for the Development of Metadata Editors, Applicability to the Annotation of Geographic Information Resources. *Data Knowl. Eng.* **2012**, *81–82*, 118–139. [CrossRef]

29. Cox, S.J.D.; Simons, B.; Car, N. Some problems with standard geospatial metadata. In Proceedings of the 8th Locate Conference, Brisbane, Australia, 10–12 March 2015. Available online: https://publications.csiro.au/ rpr/download?pid=csiro:EP149934&dsid=DS2 (accessed on 20 August 2018).

30. Shade, S.; Cox, S.J.D. Linked Data in SDI or How GML Is Not about Trees. In Proceedings of the 13th AGILE Conference, Guimarães, Portugal, 10–14 May 2010. Available online: http://agile2010.dsi.uminho.pt/pen/ ShortPapers_PDF/73_DOC.pdf (accessed on 20 August 2018).

31. Godfrey, B.; Kenyon, J. The Geospatial Metadata Manager's Toolbox: Three Techniques for Maintaining Records. *Code4Lib J.* **2015**, *29*. Available online: http://journal.code4lib.org/articles/10601 (accessed on 20 August 2018).

32. 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 20 August 2018).

33. 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?csn umber=39890 (accessed on 20 August 2018).

34. 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 20 August 2018).

35. Federal Geographic Data Committee. North American Profile of ISO 19115:2003, Geographic Information—Metadata, 2009. Available online: http://www.fgdc.gov/standards/projects/incits-l1-standards-projects/NAP-Metadata (accessed on 20 August 2018).

36. Holland, P.; Williamson, I.; Rajabifard, A.; Manning, J. Making the SDI concept relevant to Asia-Pacific countries: The PCGIAP experience. In Proceedings of the FIG Working Week/GSDI-8, Cairo, Egypt, 16–21 April 2005. Available online: https://minerva-access.unimelb.edu.au/handle/11343/33847 (accessed on 20 August 2018).

37. Australia New Zealand Land Information Council. ANZLIC Metadata Profile, 2007. Available online: http://www.anzlic.gov.au/resources/metadata (accessed on 20 August 2018).

38. Chrisman, N. *Exploring Geographic Information Systems*; Wiley: Hoboken, NJ, USA, 1997.

39. Fugazza, C.; Tagliolato, P.; Frigerio, L.; Carrara, P. Web-Scale Normalization of Geospatial Metadata Based on Semantics-Aware Data Sources. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 354. [CrossRef]

40. 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. [CrossRef]

41. W3C XML Schema Working Group. XML Schema Part 1: Structures Second Edition. W3C Recommendation 28 October 2004, World Wide Web Consortium, 2004. Available online: https://www.w3.org/TR/xmlschema-1/ (accessed on 20 August 2018).

42. AgID. Repertorio Nazionale dei Dati Territoriali—RNDT. Technical Report, Agenzia per l'Italia Digitale, 2012. Available online: http://archivio.digitpa.gov.it/repertorio-nazionale-dei-dati-territorial (accessed on 20 August 2018).

43. Fugazza, C.; Pepe, M.; Oggioni, A.; Tagliolato, P.; Carrara, P. Streamlining geospatial metadata in the Semantic Web. *IOP Conf. Ser. Earth Environ. Sci.* **2016**, *34*. [CrossRef]

44. Tagliolato, P.; Oggioni, A.; Fugazza, C.; Pepe, M.; Carrara, P. Sensor metadata blueprints and computer-aided editing for disciplined SensorML. *IOP Conf. Ser. Earth Environ. Sci.* **2016**, *34*. [CrossRef]

45. Oggioni, A.; Tagliolato, P.; Fugazza, C.; Pepe, M.; Menegon, S.; Pavesi, F.; Carrara, P. Interoperability in marine sensor networks through SWE services. In *Oceanographic and Marine Cross-Domain Data Management for Sustainable Development*; IGI Global: Hershey, PA, USA, 2016.

46. Cox, S. OWL Representations of the ISO/TC 211 Harmonized UML Model For Geographic Information, 2013. Available online: http://def.seegrid.csiro.au/static/isotc211/ (accessed on 20 August 2018).

47. World Wide Web Consortium. Data Catalog Vocabulary (DCAT), 2014. Available online: https://www.w3.org/TR/vocab-dcat/ (accessed on 20 August 2018).

48. European Commission. DCAT Application Profile for Data Portals in Europe (DCAT-AP), 2013. Available online: https://joinup.ec.europa.eu/asset/dcat_application_profile/ (accessed on 20 August 2018).

49. European Commission. GeoDCAT-AP: A Geospatial Extension for the DCAT Application Profile for Data Portals In Europe, 2015. Available online: https://joinup.ec.europa.eu/node/148281 (accessed on 20 August 2018).

50. Fugazza, C.; Vaccari, L. Coupling human- and machine-driven mapping of SKOS thesauri. *Int. J. Metadata Semant. Ontol.* **2011**, *6*, 155. [CrossRef]

51. Reed, C.; Botts, M.; Davidson, J. OGC® sensor web enablement: overview and high level architecture. In Proceedings of the IEEE Autotestcon, Baltimore, MD, USA, 17–20 September 2007; pp. 372–380. [CrossRef]

52. 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. Available online: http://www.igi-global.com/chapter/methodologies-augmented-discovery-geospatial-resources/70448 (accessed on 20 August 2018).

53. Fugazza, C. Toward semantics-aware annotation and retrieval of spatial data. *Earth Sci. Inf.* **2011**, *4*, 225–239. [CrossRef]

54. Verborgh, R.; Sande, M.V.; Hartig, O.; Herwegen, J.V.; Vocht, L.D.; Meester, B.D.; Haesendonck, G.; Colpaert, P. Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. *Web Semant. Sci. Serv. Agents World Wide Web* **2016**, *37–38*, 184–206. [CrossRef]

55. Fionda, V.; Pirrò, G.; Consens, M.P. Extended Property Paths: Writing More SPARQL Queries in a Succinct Way. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15), Austin, TX, USA, 25–30 January 2015; pp. 102–108. Available online: http://dl.acm.org/citation.cfm?id=2887007.2887022 (accessed on 20 August 2018).

56. W3C—Web Annotation Working Group . Web Annotation Data Model, 2017. Available online: https://www.w3.org/TR/annotation-model/ (accessed on 20 August 2018).