International Journal on Advances in Intelligent Systems, vol 3 no 3 & 4, year 2010, http://www.iariajournals.org/intelligent_systems/

Aggregating Geoprocessing Services using the OAI-ORE Data Model

Carlos Abargues, Carlos Granell, Laura Díaz, Joaquín Huerta
Institute of New Imaging Tecnologies (INIT)
Universitat Jaume I (UJI)
Castellón, Spain
{*abargues, carlos.granell, laura.diaz, huerta*}*@uji.es*

***Abstract*—Rapid discovery and access of geospatial resources is critical for many application domains that require agile data integration. In this context, cross-domain geospatial applications need immediate access to geospatial resources of interest in order to rapidly integrate them in scalable, functional Web applications. In this paper we explore new perspectives to build pragmatic geospatial Web applications, drawing on the ideas of recent initiatives like Linked Open Data and Open Archives Initiative. By using and extending standards and principles from these initiatives we are able to model single and composite geoprocessing services as a collection of heterogeneous Web resources. Such collections are built by the principle aggregation by linking that enables to connect and link multiple geospatial data and services across different application domains.**

***Keywords*-models for geocomputation; geoprocessing services; collections; service integration; OAI-ORE; Linked Open Data**

I. Introduction

Geospatial data sets are increasingly becoming available in open repositories. Not only as official, validated data sets collected by authorities and experts that make them available through catalogues in Spatial Data Infrastructures (SDI) nodes, but also as on-line resources dispersed everywhere produced by thousands of individuals. Nonexpert users are taking the role of productors of geospatial data through the massive use of social networks (Flickr, Twitter, etc.) and location-based devices (mobile phones, digital cameras, etc.), which leads to huge amounts of rich georeferenced user-generated content in a great variety of sizes and formats [2]. As some authors pointed out [3], SDI nodes cannot ignore the millions of users providing up-to-date data anywhere at any time, becoming a challenging task for the next generation of geospatial applications to conciliate the up-bottom approach in traditional SDI creation with the bottom-up approach powered by users [4].

In the SDI context geoprocessing services are a powerful means for creating web-based applications, integrating geospatial data from multiple sources [5] [6] [7]. In addition, Brauner et al. [8] have recently reported a set of research projects focused on designing and implementing geoprocessing services, in order to draw a research agenda for future developments in the realm of distributed geoprocessing computing. According to these studies [3] [8], some issues that should be addressed in the near future are:

- The need of integrating information from multiple and heterogeneous sources to benefit from the rich, valuable, up-to-date user-generated content.
- The lack of flexible mechanisms to create on-demand, scalable service-oriented geospatial solutions that contains relevant geospatial resources such as geoprocessing services, geospatial data and user-generated content.

Our long-term research goal pursues new perspectives to provide pragmatic, scalable approaches to creation of geoprocessing workflows that take into account user-generated content out of SDI context. Here we focus on the facet of description (leaving other facets like execution aside for now) of geoprocessing services in such a way that let users create collections of related geoprocessing services.

The proposed approach draws on the principles and best practices exposed by Linked Open Data (LOD) [9] [10]. This initiative stems from the very principles of Web architecture and pursues the goal of “enabling people to share structured data on the Web as easily as they can share currently Web resources” [9]. The idea consists of publishing data in a structured manner and creating typed links between data resources from heterogeneous sources, assuming these two tenets: (i) RDF (Resource Description Framework) data model is used as common model to publish structured data; and (ii) extensive use of typed links to connect data from different data sources. Essentially, it is assumed that the more interlinked data, the more aspects of meaning (richness) might be represented.

Although accessing to data itself may be viewed as a prior goal, users are already retrieving georeferenced data either through SDI data download services or from open repositories such as Flickr, Open Street Maps and Twitter. The availability and stability of latter data sets are fundamental and should be reinforced, though, the real impediment is to provide better connections (interlinking) among the vast amount of user-generated data, SDI content, and geoprocessing services so that experts and nonexperts alike may access to structured data and services for their cross-domain geospatial applications. Furthermore, our aim here is to apply the OAI-ORE (Open Archive Initiative -

Object Reuse and Exchange) protocol (see Section II-C), which has been proven to be successful in digital library projects, extending it with new types of relationships to the geospatial landscape in order to support not only structured data, as LOD states, but also interlinking other types of Web resources like geoprocessing services.

This paper is an extended version of a conference paper referenced in [1]. The rest of the paper is structured as follows. In Section II we introduce the basic concepts used throughout this paper focusing specifically on the OAI-ORE's abstract data model. The proposed approach to model geoprocessing services and data as collections of interlinked heterogeneous resources is described in Section III. Section IV compares our work with related projects. Finally, Section V concludes by summarizing the key features of our approach and discussing ongoing work.

II. Background

This section presents relevant concepts and definitions for the remaining sections. In particular, the following subsections will shortly introduce the SDI and related geospatial services, the LOD project, and the OAI-ORE specification.

A. SDI and geoprocessing services

Spatial Data Infrastructures (SDI) describe the notion of service-oriented management, accessing, and processing of geospatial data. The implementation of SDI has traditionally followed a service oriented architecture paradigm where web services technology plays a enabling role in data integration and promotion of interoperability among heterogeneous distributed information sources [11].

Geospatial web services allow users to access, manage, and process geospatial data in a distributed manner [12]. The demand for interoperability has boosted the development of standards and tools to facilitate data transformation and integration, mostly in terms of standard interfaces specified by Open Geospatial Consortium (OGC[1]) and Technical Committee 211 (TC211[2]) of International Organization for Standardization (ISO). The Web Map Service (WMS), the Web Feature Service (WFS) and the Web Coverage Service (WCS) are some prominent examples of OGC interfaces for geospatial services. All come in different versions, where WMS 1.3.0 [13], WFS 1.1.0 [14], and WCS 1.1.2 [15] are the most recent. The central building-blocks for data, as well as service discovery, are provided by the Catalogue Services for the Web (CSW) [16] and so called geoportals [17]. The CSW provides one access point to users that search for geospatial data.

Geoprocessing services essentially transform geospatial data to produce new data or meaningful information [18]. A substantial leap ahead in the domain of geoprocessing services was the OGC Web Processing Service (WPS) specification [19]. This specification was designed to encapsulate generic operations and algorithms over the Internet. The basic operational unit of the OGC WPS is the notion of process, that is, a geospatial operation with inputs and outputs of a defined type. This means that a given WPS instance (a concrete WPS service running) may offer one or various operations (or processes) as normal web services do. The common communication pattern between a client and a WPS instance encompasses three types of requests. A request can be sent to the WPS instance via HTTP GET with parameters provided as Key-Value Pairs (KVP) or via HTTP POST, with parameters supplied in a XML document. These three types of requests are:

- *GetCapabilities*. First, a WPS instance receives a KVP *getCapabilities* request (which is common for all OGC geospatial services) and simply responds with an XML document, containing metadata such as server provider, contact information, general description, and a list of contained geoprocessing operations (processes) offered by the queried WPS instance.
- *DescribeProcess*. A WPS-client selects a process identifier from the *getCapabilities* response and performs a *describeProcess* request, either as KVP or as XML document. The WPS instance responds with an XML document containing needed information for the solicited process, such as input and output parameter names and types, so that the WPS-client may later build the execute request.
- *Execute*. The WPS-client eventually requests the execution of a geospatial operation, with all required input data by invoking the execute method as an XML document request. The WPS instance then runs the operation and returns the results informing also of its status.

As geospatial web solutions continue to grow and increase in complexity, many standards organizations, industry bodies, and the geospatial research community have paid attention to the effective composition and orchestration of geospatial web services [20]. Rather than describing a new service interface for geoprocessing services, our aim here is to propose a new way to improve service compositions in terms of collections or aggregations of geoprocessing services in line with the principles of the Linked Data community.

B. Linked Open Data

Linked Open Data (LOD) represents a style of information publishing on the Web. This style relies on traditional web technologies and the usage of light-weight techniques for data model representation. The former resides on the use of Uniform Resource Identifiers (URI) as reference points. A URI is used to uniquely identify a resource, i.e., a piece of data, and also for actual access to the resource

[1] http://www.opengeospatial.org/
[2] http://www.isotc211.org/

representation. This implies that HTTP URI should be deferenciable URI, that is, user can look up these URIs to retrieve resource representations. Content negotiation comes here to allow clients to specify an acceptable representation of a data set [21]. While connecting to a data source, the client may specify the desired representation. This may be, for instance, plain RDF, or an HTML representation with increased readability for the human user.

The former refers to the Resource Description Framework (RDF) as basic structure for any form of description. RDF provides means to describe any kind of resource in form of triples (subject-predicate-object). In this way, data published according to LOD principles is exposed in RDF format and interlinked by exploting the intrinsic capabilities of the RDF model to link resources. As we will see in the next section, OAI-ORE and LOD share some characteristics what make OAI-ORE a suitable candidate to model collections of geoprocessing services with the benefits of the LOD project.

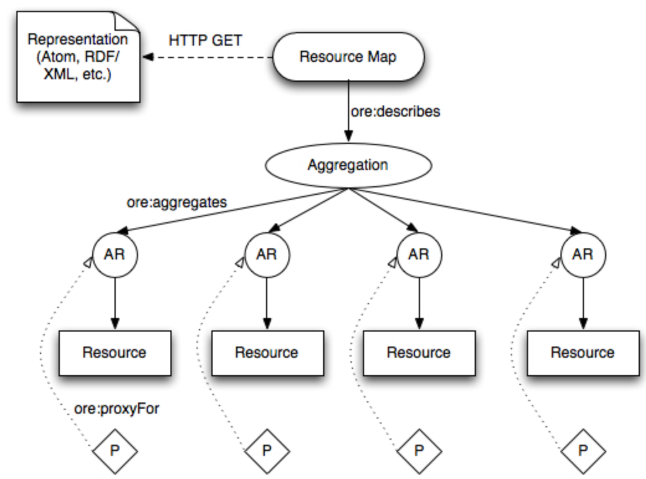

Figure 1. Simplified OAI-ORE's abstract data model

C. OAI-ORE's abstract data model

The Open Archive Initiative - Object Reuse and Exchange (OAI-ORE) protocol [22] defines an abstract data model [23] for describing, reusing, and exchanging collections of Web resources. The aim of this protocol is to expose rich content (text, images, data, video, etc.) in aggregations to be then fed by applications in the realm of digital library domain. Obviously, OAI-ORE is closely related with the OAI - Protocol for Metadata Harvesting (OAI-PMH) [24], since for instance source content (e.g., e-prints records) described in OAI-ORE can be harvested automatically in order to replicate it in others remote repositories [25].

Conceptually, the OAI-ORE's abstract data model builds strongly on the principles defined on LOD. First, the notion of "addressable resources" indicates that resources of any type (file, image, text document, metadata, process, etc.) should be identified using HTTP URI. Secondly, exploiting the simple mechanism of "typed links" to connect resources enables the discovery, browsing, and access to more related and connected data.

In principle, the use of the capabilities of OAI-ORE and LOD to build scalable, distributed Web applications that integrate heterogeneous remote sources becomes evident [26]; LOD takes the principles of the current Web architecture, the most, by far, scalable and distributed information system. As the OAI-ORE's abstract data model relies on such principles, we find a rational argument to use it to link collections of Web resources, considering a geoprocessing service as a particular type of Web resource.

Before discussing how the OAI-ORE's abstract data model can be used for describing and linking geoprocessing services (see Section III), we introduce here its key entities [23]. The simplified diagram in Figure 1 shows how these entities are related each other. The entity Aggregation plays a central role as it represents a collection of addressable resources that in turn are called Aggregated Resources (AR). The ore:aggregates relation denotes here the "aggregation by linking" mechanism to connect resources related in some way. This implies that both Aggregations and AR entities are addressable resources in the sense that both use HTTP URI as referencing method. Here comes the process of dereferencing URI, another key aspect in LOD, which just means looking up a URI on the Web in order to get either the resource itself or its representation.

Aggregation and AR are abstract terms that must still refer to concrete resources, which can be of any type such as a document, image, process, service, and even a chain of geoprocessing services as we will see in the following section. The OAI-ORE specification makes use of the Resource Map entity to provide a concrete representation for the whole aggregation, mostly derived from RDF. Some suggested formats in the specification are the Atom syndication format [27], RDF/XML[3], and RDFa[4] (a microformat for extending XHTML to support RDF). We use the RDF/XML serialization so that resulting collections can be readily added and connected to other LOD datasets since RDF triples are the common data model.

OAI-ORE defines a useful abstract entity called Proxy (P) by which it is possible to express the role an aggregated resource has explicitly in the context of an aggregation. For example, two resources may have a temporal relationship that connects to each other and this is only meaningful within the aggregation context in which they are defined. The use of relationships from and to Proxy elements instead of the Aggregated Resource elements they represent does not affect the original resource. In addtion, the use of Proxy elements

[3]http://www.w3.org/TR/rdf-syntax-grammar/
[4]http://www.w3.org/TR/rdfa-syntax/

does not make private information explicit (e.g., temporal and semantic relationships) to external aggregations because either this knowledge is not required or is meaningless to them. The Proxy entity thus enables encapsulation, a major driving force to support reusability [28].

Besides the Proxy entity, OAI-ORE protocol permits the use of relationships to link directly to resources and aggregations. These relationships can be either internal, between the resources defined within the aggregation, or external, linking to external resources such as georeferenced user-generated content. In both cases, specific relationships such as ore:aggregates and ore:describes are defined. An example of external relationship is ore:similarTo. However, following the "typed link" principle of LOD, those resources defined in a given OAI-ORE aggregation may, indeed should, link to and be linked from other external resources based on relationships characterized semantically by other vocabularies.

Aside from linking related resources, the "aggregation by linking" mechanism makes it ease to create complex hierarchical aggregations from simpler ones. This implies that collections of resources can be scaled and reused easily, since incorporating or eliminating a resource from a given collection simply means to refer or not to its HTTP URI. Next section focuses on how geoprocessing services can be seen as resources and described using the OAI-ORE's abstract data model.

III. Approach

This section first proposes the conceptual architecture that supports our approach and lists some assumptions that drive our research at the present stage. Then a set of new relationships to model geoprocessing services inside an OAI-ORE colection is presented as an extension on this protocol. Finally we describe how to use the OAI-ORE and the new extension to model geoprocessing service and their composition.

A. Architecture

SDI-based applications are built upon a multilayer architecture as depicted in the right side of Figure 2 (blue boxes). Application layer may contain thin client tools such as geoportals, mashups and rich internet applications (e.g., Flex, JavaFX), and also thick desktop-based clients. The middleware layer comprises multiple distributed services, which allow client applications to discover, access, and process geospatial data and metadata from remote repositories (Data layer). SDI applications work in this way because SDI nodes are normally architectured in such a way[5].

To provide better connections with other type of data out of SDI content, like user-generated content and LOD datasets (blue clouds and gray circles in Figure 2), we have added a Transformation process to convert heterogeneous

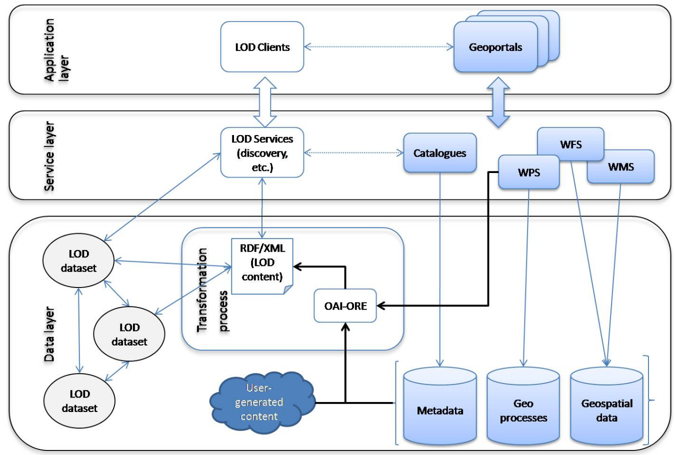

Figure 2. Proposed architecture

source resources into RDF/XML format. The source resources may be geospatial data sets in SDI repositories, georeferenced documents, pictures or tweets (Twitter messages), as well as geoprocessing services to compute calculations over such datasets. The black arrows in Figure 2 shows the two-step transformation process: first resource collections are generated based on the OAI-ORE data model to be then serialized as RDF/XML documents. Besides resulting collections may be readily connected to resources in other RDF-based LOD datasets[6] across several communities and domains [9].

The proposed architecture poses some requirements for interlinking geoprocessing resources: addressable resources (data and services), connecting (linking) resources, and the ability to see and link resource descriptions (metadata) for each resource. In the following we enumerate some assumptions taken in this paper for describing collections of interlinked Web resources:

- Geospatial datasets and services are addresable resources, i.e., have referenciable HTTP URI [9] [10] [26].
- A geoprocessing service is considered a type of Web resource, which let us rely on the OAI-ORE protocol to model interlinking geoprocessing services.
- Since a geoprocesing service is a resource, a chain of geoprocessing services may be then comparable to a collection of heterogeneous Web resources. A collection thereby is a resource with own metadata.

B. Extending the OAI-ORE abstract data model

A geoprocessing service can be characterized by its signature: a function or capability, and a set of input and output parameters. Taking this simple approach, the mapping to the OAI-ORE's abstract data model is driven by the following

[5]http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3_5_INSPIRE_NS_Architecture_v3-0.pdf

[6]http://richard.cyganiak.de/2007/10/lod/

rules: (i) a geoprocessing service is an Aggregation entity; (ii) its components are Aggregated Resources, i.e., the capability of the resource, and the collections of input and output parameters; (iii) a Resource Map entity then describes the Aggregation using a serialization format.

These assumptions describe a method for mapping geoprocessing service elements with the OAI-ORE data model, however, the built-in relationships defined by this specification are not enough to express the roles and relationships for modelling aggregations of geoprocessing services. In order to appropiately define these relationships, we have extended the OAI-ORE abstract data model with a set of new relationships (with prefix *ores*) that can be applied to the already existing elements.

Table I briefly shows these new relationships altogheter with their name, URI, the inverse relationship in the case this exists and their domain and range properties. Not only the definition of these relationships but also the elements that make use of them is an important aspect to consider. The relationships among the different elements that describe a geoprocessing service can only be meaningful in the context of this description and they may not be useful or representative outside the geoprocessing service description for an aggregated resource. In order to encapsulate this information most of the new relationships defined are applied to the Proxy element. The Proxy actually becomes a key element for describing the geoprocessing service representing and containing the internal relationships that models it.

Geoprocessing services may receive an input to generate an output or result. This behaviour is modeled in our approach through the use of the relationships ores:aggregatesInput, ores:inputAggregatedBy, ores:aggregatesOutput and ores:outputAggregatedBy. The first two relationships allow the specification of the input for a given service being each relationship the inverse for the other. The last two ones do the proper in the task of specifying the output for the service representing again an inverse relationship among them. Since a service can have zero or more inputs (and outputs), the cardinality of these relationships can be specified as zero to any.

Similar behaviour occurs when defining the inner processes that compounds the geoprocessing service. This components can also specify their inputs and outputs using the declared relationships ores:inputFor, ores:hasInput, ores:outputFor, ores:hasOutput. A given process may have zero or more inputs and zero or more outputs that can be specified by using any number of relationships to link both, the process and the inputs or outputs represented by their corresponding Proxy element.

The different processes that compound the geoprocessing service have to be described in a certain partial order. To model this feature, the proposed extension of the OAI-ORE data model defines the relationships ores:next and ores:previous. These two new relationships can be applied only to those Proxy elements that represent processes as aggregated resources and indicate the next or the previous processes that should be followed during the service execution. Thanks to these relationships it is possible not only to specify an order for the inner processes but also navigate among them and check for instance the requirements (i.e., output of a process that serves as input for another) in order to start the execution of a process.

OAI-ORE defines a method for reusing defined aggrega-

Table I
OAI-ORE'S MODEL EXTENSION FOR SERVICE DESCRIPTION

Name	ores:**aggregatesInput**
URI	http://www.geoinfo.uji.es/ores/terms/aggregatesInput
Inverse Of	ores:inputAggregatedBy
Domain	ore:Aggregation
Range	ore:Proxy
Name	ores:**inputAggregatedBy**
URI	http://www.geoinfo.uji.es/ores/terms/inputAggregatedBy
Inverse Of	ores:aggregatesInput
Domain	ore:Proxy
Range	ore:Aggregation
Name	ores:**aggregatesOutput**
URI	http://www.geoinfo.uji.es/ores/terms/aggregatesOutput
Inverse Of	ores:outputAggregatedBy
Domain	ore:Aggregation
Range	ore:Proxy
Name	ores:**outputAggregatedBy**
URI	http://www.geoinfo.uji.es/ores/terms/outputAggregatedBy
Inverse Of	ores:aggregatesOutput
Domain	ore:Proxy
Range	ore:Aggregation
Name	ores:**inputFor**
URI	http://www.geoinfo.uji.es/ores/terms/inputFor
Inverse Of	ores:hasInput
Domain	ore:Proxy
Range	ore:Proxy
Name	ores:**hasInput**
URI	http://www.geoinfo.uji.es/ores/terms/hasInput
Inverse Of	ores:InputFor
Domain	ore:Proxy
Range	ore:Proxy
Name	ores:**outputFor**
URI	http://www.geoinfo.uji.es/ores/terms/outputFor
Inverse Of	ores:hasOutput
Domain	ore:Proxy
Range	ore:Proxy
Name	ores:**hasOutput**
URI	http://www.geoinfo.uji.es/ores/terms/hasOutput
Inverse Of	ores:outputFor
Domain	ore:Proxy
Range	ore:Proxy
Name	ores:**next**
URI	http://www.geoinfo.uji.es/ores/terms/next
Inverse Of	ores:previous
Domain	ore:Proxy
Range	ore:Proxy
Name	ores:**previous**
URI	http://www.geoinfo.uji.es/ores/terms/previous
Inverse Of	ores:next
Domain	ore:Proxy
Range	ore:Proxy

tions in others as aggregated resources of the latter indicating the resource map or serialization of the former by the relationship ore:isDescribedBy that links both. The same may happen when describing a geoprocessing service, however, in this case the aggregations can have two different roles: either describing a collection of aggregated resources or describing a executable geoprocessing service. In the first case all of the aggregated resources could be used as input for a given process. In the second case, through, it makes sense that only the aggregated resources that represent the output of its execution may be of interest instead of the entire service description (collection).

This behaviour represents an ambiguity concerning the role a nested aggregation plays in a service description. To resolve it the relationship ores:hasOutput can be used to specify in both cases those elements that can be reused by other geoprocessing service descriptions when the first aggregation is used as one of its aggregated resources. The use of ores:hasOutput makes it easy to reuse the output of a geoprocessing service or a collection of other resources and also allows to specify only a subset of the aggregated resources in a collection or intermediate results in a geoprocessing service that may be of interest although they are not the result for the service execution.

C. Modeling geoprocessing services as Web resources

Considering the OAI-ORE specification and the previously explained extension based on it, let us consider now a concrete geoprocessing service like a transformation service that converts a source KML (Keyhole Markup Language) file [29] into a GML (Geography Markup Language) format [30]. This service is called Kml2Gml and takes one input parameter –a data resource–, and returns the corresponding GML content as an addressable resource so that can be retrieved by dereferencing its URI.

Figure 3 illustrates this simple scenario using a named graph to represent the mappings between the Kml2Gml service and the OAI-ORE's abstract data model including the relationships defined in the extension for describing geoprocessing services. From top to bottom, the Resource Map entity named ReM-1 describes, through the relationship ore:describes, the Aggregation entity A-1. The A-1 entity is composed, through the relationship ore:aggregates, of three Aggregated Resource entities, named AR-1, AR-2, and AR-3 respectively. These AR entities map to counterpart resources of the Gml2Kml service. In this particular case, AR-2 represents the function, AR-1 the input resource, and AR-3 the output resource. Their corresponding Proxy elements are also represented in the graph by the elements P-1, P-2 and P-3 respectevily and allow the definition of the required relationships for the aggregated resources that are meaningful only in the service description context. Table II lists the set of URL for the resulting addressable resources, both abstract (Aggregation, Aggregated Resource and Proxy entities) and concrete resources such as geospatial data files and services.

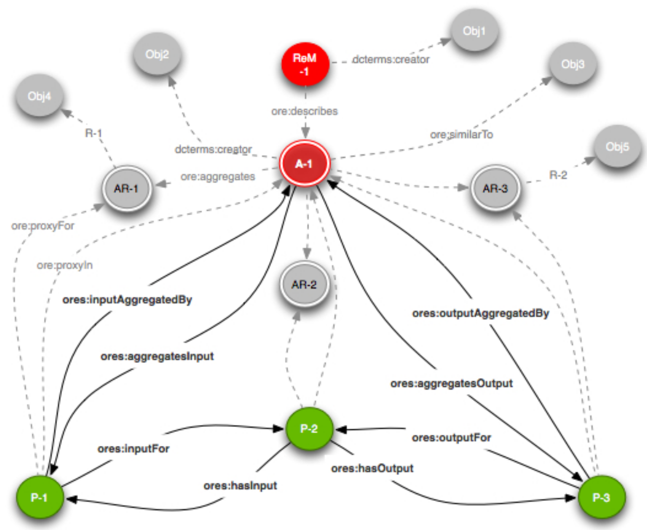

Figure 3. Mapping a geoprocessing service in OAI-ORE

Table II
LIST OF ADDRESSABLE RESOURCES

ENTITIES	URI
ReM-1	http://www.geoinfo.uji.es/resource/aggregation.rdf
A-1	http://www.geoinfo.uji.es/resource/aggregation
AR-1	http://www.geoinfo.uji.es/data/datasetKML.kml
AR-2	http://www.geoinfo.uji.es/process/Kml2Gml
AR-3	http://www.geoinfo.uji.es/data/datasetGML.gml
P-1	http://www.geoinfo.uji.es/proxy/r? what=http://www.geoinfo.uji.es/data/datasetKML.kml& where=http://www.geoinfo.uji.es/resource/aggregation
P-2	http://www.geoinfo.uji.es/proxy/r? what=http://www.geoinfo.uji.es/process/Kml2Gml& where=http://www.geoinfo.uji.es/resource/aggregation
P-3	http://www.geoinfo.uji.es/proxy/r? what=http://www.geoinfo.uji.es/data/datasetGML.gml& where=http://www.geoinfo.uji.es/resource/aggregation

Figure 3 shows in grey tone all those relationships not defined by the presented extension. These relationships include those externally defined such as dcterms:creator, used by the Resource Map and Aggregation entities to point to an author (external) resource based on the Dublin Core vocabulary[7].

As described previously, AR-1 and AR-3 entities represent the needed input and output resources for the capability resource AR-2. To capture these relationships between resources in the OAI-ORE's abstract data model, it is required a couple of steps. First, we create a Proxy entity for each AR entity (P1, P2, and P3 respectively). So a Proxy entity is tied to the corresponding AR entity by two OAI-ORE built-in relationships, namely ore:proxyFor and ore:proxyIn. The former indicates that the Proxy entity is for a concrete

[7] http://dublincore.org/documents/dces/

Table III
LIST OF ADDRESSABLE RELATIONS

RELATIONS	URI
ore:similarTo	http://www.openarchives.org/ore/terms/similarTo
ore:describes	http://www.openarchives.org/ore/terms/describes
ore:aggregates	http://www.openarchives.org/ore/terms/aggregates
ore:proxyFor	http://www.openarchives.org/ore/terms/proxyFor
ore:proxyIn	http://www.openarchives.org/ore/terms/proxyIn
dcterms:creator	http://purl.org/dc/terms/creator
ores: inputAggregatedBy	http://www.geoinfo.uji.es/ores/terms/ inputAggregatedBy
ores: aggregatesInput	http://www.geoinfo.uji.es/ores/terms/ aggregatesInput
ores: outputAggregatedBy	http://www.geoinfo.uji.es/ores/terms/ outputAggregatedBy
ores: aggregatesOutput	http://www.geoinfo.uji.es/ores/terms/ aggregatesOutput
ores:inputFor	http://www.geoinfo.uji.es/ores/terms/inputFor
ores:hasInput	http://www.geoinfo.uji.es/ores/terms/hasInput
ores:outputFor	http://www.geoinfo.uji.es/ores/terms/outputFor
ores:hasOutput	http://www.geoinfo.uji.es/ores/terms/hasOutput

AR entity. The latter is informative and just keeps informed the upper Aggregation A-1 about the active Proxy entities. Second, semantic relationships among resources are possible by means of relationships belonging to vocabularies, either in the OAI-ORE domain itself or in others.

It is here where the extension plays its role modelling the geoprocessing service through the use of the requiered relationships. For example ores:hasInput and ores:inputFor indicate that AR-1 (input resource) serves as input to the AR-2 (capability resource). Something similar happens for the ores:hasOutput and ores:outputFor relationship but involving the AR-2 and AR-3 entities to indicate the output for the former. The relationships ores:inputAggregatedBy and its invers ores:aggregatesInput indicate that the entity AR-1 through its Proxy entity P-1 serve as input for the aggregation A-1 that models the entire service. In order to indicate the output of the service the relationships ores:outputAggregatedBy and ores:aggregatesOutput links both the Aggregation A-1 and the Aggregated Resource AR-3 through its Proxy P-3.

Although OAI-ORE data model and the proposed extension define the required relationships to model geoprocessing services as aggregations, other relationships and terms can be used to enrich a description of a collection. These terms and relationships can be created for a concrete purpose or reused from common vocabularies in other fields. The use of common vocabularies in the geoprocessing context is a tricky question. Some authors claim the need of a geoprocessing taxonomy [8], as a mechanims to provide a common semantic background on which discovery, access, and composition of geoprocessing services can be achieved. In other words, for effective interoperability it is a must to build and label similar resources in compatible ways. Otherwise, the task of choosing, extending, and merging vocabularies becomes a sensitive issue [31].

As a best practice, semantic terms should be reused from well-known vocabularies wherever possible, avoiding the definition of new terms if they already exist. As OAI-ORE protocol supports the use of existing vocabularies, we have followed this strategy by reusing relationships from well-known vocabularies like RDF schemas[8] (e.g., rdf:domain). In this way, additional metadata for the whole collection (e.g., use case, context, data provenance) may also be encoded as external resources using meaningful relationships. Table III lists the most relevant relationships used in the Km2Gml geoprocessing service scenario.

Another example of widely-used vocabulary with geographic connotations is the case of the vocabulary or ontology offered by Geonames[9]. This vocabulary is expressed in OWL [32] and offers a collection of over six million of place names and other relevant terms to express relationships. In addition, Geonames features are interlinked each other by means of typed links denoting hierarchical inclusion (e.g., continent, countries, administratives units, etc.) and proximity. The Geonames vocabulary seems to be a feasible choice to link geospatial resources (user-generated content, SDI content, OAI-ORE aggregations), thus offering a simple mechanism of georeferencing heterogeneous Web resources. Therefore, the combination of Proxy entities and meaningful relationships from (existing) vocabularies enables the definition of flexible and customized connections among disparate resources of a given aggregation.

Finally the code in Figure 4 shows a portion of a possible RDF/XML representation that describes a geoprocessing service as represented in Figure 3 following the OAI-ORE recommendations for serializing aggregations.

D. Interlinking geoprocessing services

The previous section described how OAI-ORE's abstract data model is used and extended to describe a geoprocessing service. To illustrate our approach a simple geoprocessing services was used as example of aggregation of resources. Independently of its complexity any geoprocessing service can be modeled following the same pattern: a capacity resource, one or more input resources and one or more output resources. This section sketches how a chain of geoprocessing services is modelled as a collection of interlinked aggregated resources and how another collection can be reused as input for a capacity resource.

Figure 5 shows a chain of two services represented in OAI-ORE. In this case, the AR-2 and AR-4 entities represent geoprocessing services. Returning to our simple scenario, AR-2 refers to the Kml2Gml service while the AR-4 would be a topology function like intersection that operates over the results of AR-2 and a collection of geometries defined in AR-5. Again, the key aspect is the combination of Proxy entities and suitable relationships that express properties

[8]http://www.w3.org/TR/rdf-schema/
[9]http://www.geonames.org/ontology

```xml
<rdf:Description rdf:about="http://www.geoinfo.uji.es/resource/aggregation.rdf">
  <rdf:type rdf:resource="http://www.openarchives.org/ore/terms/ResourceMap"/>
  <ore:describes rdf:resource="http://www.geoinfo.uji.es/resource/aggregation"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/resource/aggregation">
  <rdf:type rdf:resource="http://www.openarchives.org/ore/terms/Aggregation"/>
  <ore:similarTo rdf:resource="http://www.geoinfo.uji.es/aggregations/aggr2"/>
  <dc:creator rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    http://www.geoinfo.uji.es/personal/cabargues
  </dc:creator>
  <ore:isDescribedBy rdf:resource="http://www.geoinfo.uji.es/resource/
    aggregation.rdf"/>
  <ore:aggregates rdf:resource="http://www.geoinfo.uji.es/data/datasetKML.kml"/>
  <ore:aggregates rdf:resource="http://www.geoinfo.uji.es/process/Kml2Gml"/>
  <ore:aggregates rdf:resource="http://www.geoinfo.uji.es/data/datasetGML.gml"/>
  <ores:AggregatesInput rdf:resource="http://www.geoinfo.uji.es/proxy/r?
    what=http://www.geoinfo.uji.es/data/datasetKML.kml&
    where=http://www.geoinfo.uji.es/resource/aggregation"/>
  <ores:AggregatesOutput rdf:resource="http://www.geoinfo.uji.es/proxy/r?
    what=http://www.geoinfo.uji.es/data/datasetGML.gml&
    where=http://www.geoinfo.uji.es/resource/aggregation"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/data/datasetKML.kml">
  <ore:isAggregatedBy rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
    http://www.geoinfo.uji.es/resource/aggregation
  </ore:isAggregatedBy>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/data/datasetGML.gml">
  <ore:isAggregatedBy rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
    http://www.geoinfo.uji.es/resource/aggregation
  </ore:isAggregatedBy>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/process/Kml2Gml">
  <ore:isAggregatedBy rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
    http://www.geoinfo.uji.es/resource/aggregation
  </ore:isAggregatedBy>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/proxy/r?
  what=http://www.geoinfo.uji.es/data/datasetKML.kml&
  where=http://www.geoinfo.uji.es/resource/aggregation">
  <ore:proxyIn rdf:resource="http://www.geoinfo.uji.es/resource/aggregation"/>
  <ore:proxyFor rdf:resource="http://www.geoinfo.uji.es/data/datasetKML.kml"/>
  <ores:InputFor rdf:resource="http://www.geoinfo.uji.es/proxy/r?
    what=http://www.geoinfo.uji.es/process/Kml2Gml&
    where=http://www.geoinfo.uji.es/resource/aggregation"/>
  <ores:InputAggregatedBy rdf:resource="http://www.geoinfo.uji.es/resource/
    aggregation"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/proxy/r?
  what=http://www.geoinfo.uji.es/process/Kml2Gml&
  where=http://www.geoinfo.uji.es/resource/aggregation"/>
  <ore:proxyIn rdf:resource="http://www.geoinfo.uji.es/resource/aggregation"/>
  <ore:proxyFor rdf:resource="http://www.geoinfo.uji.es/process/Kml2Gml"/>
<ores:hasInput rdf:resource="http://www.geoinfo.uji.es/proxy/r?
  what=http://www.geoinfo.uji.es/data/datasetKML.kml&
  where=http://www.geoinfo.uji.es/resource/aggregation"/>
<ores:hasOutput rdf:resource="http://www.geoinfo.uji.es/proxy/r?
  what=http://www.geoinfo.uji.es/data/datasetGML.gml&
  where=http://www.geoinfo.uji.es/resource/aggregation"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.geoinfo.uji.es/proxy/r?
  what=http://www.geoinfo.uji.es/data/datasetGML.gml&
  where=http://www.geoinfo.uji.es/resource/aggregation">
  <ore:proxyIn rdf:resource="http://www.geoinfo.uji.es/resource/aggregation"/>
  <ore:proxyFor rdf:resource="http://www.geoinfo.uji.es/data/datasetGML.gml"/>
  <ores:OutputFor rdf:datatype="http://www.geoinfo.uji.es/proxy/r?
    what=http://www.geoinfo.uji.es/process/Kml2Gml&
    where=http://www.geoinfo.uji.es/resource/aggregation"/>
  <ores:OutputAggregatedBy rdf:resource="http://www.geoinfo.uji.es/resource/
    aggregation"/>
</rdf:Description>
```

Figure 4. RDF/XML based representation for a geoprocessing service description

over the AR entities, such as the order in which the services should be organized and what resources acts as input and output parameters.

As observed in Figure 5, the ores:next and ores:previous relationships introduce the partial order among the resources AR-2 and AR-4 of the collection. Also, the ores:hasInput,

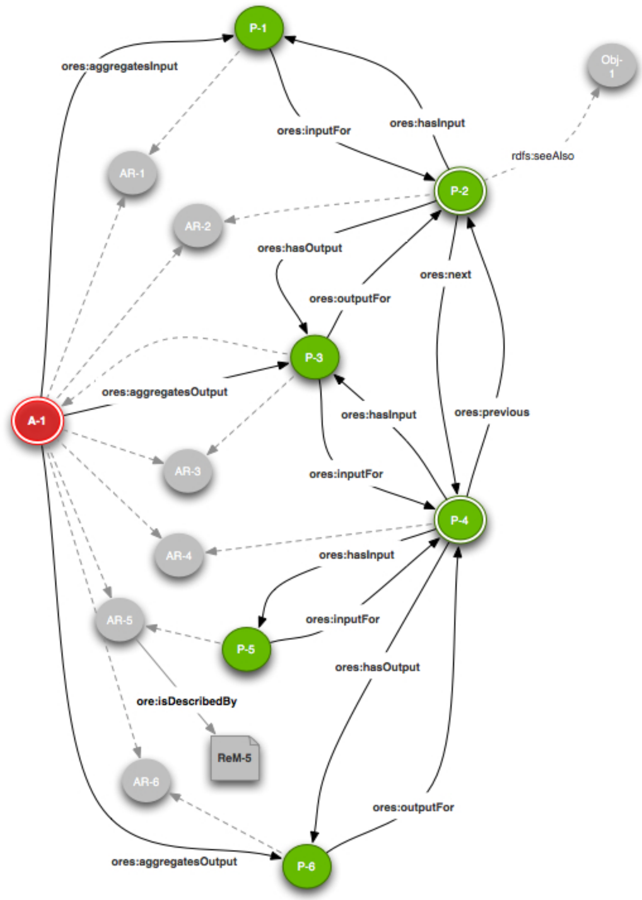

Figure 5. Interlinking geoprocessing services in OAI-ORE

ores:hasOutput and their inverse relationships connect properly all of the aggregated resources. Note that the Proxy entity P-3 has two relations –ores:inputFor and ores:outputFor–, but each one refers to disjoint Proxy entities (P-2 and P-4 respectively). This means that the entity P-3 plays a different role depending on which target entity is connected to.

In this scenario the geoprocessing service referenced by the resource AR-4 and its Proxy entity P-4 represents an intersection function that accepts two inputs: a GML geometry returned by the Kml2Gml service and a second geometry defined by the resource AR-5. The latter represents in fact another collection as indicated by the relationship ore:isDescribedBy that links the Aggregated Resource entity with the Resource Map entity for serialization. As indicated in previous sections, the taks of nesting and reusing aggregations involves the use of OAI-ORE built-in relationships along with the proposed ores:aggregatesOutput (and/or its inverse relationship ores:outputAggregatedBy when required) for indicating those resources that can be reused in other collections. By using these relationships all the geometries required to perform the intersection function through the capacity resource AR-4 are indicated conveniently in the collection AR-5. Similarly the different outputs generated by the execution of the interlinked geoprocessing services in our example are anotated by using these relationships. It is important to note that not only the final result is indicated (intersection function result) but also intermediate results (Kml conversion to Gml).

Connecting to alternative resources is also possible and even desirable. The P-2's outcomming relationship rdfs:seeAlso lets us connect to other functional-like services for data transformation. In practice, therefore, composing and reusing entire aggregations and aggregated resources by "aggregation by linking" mechanism is possible and even encouraged in the specification of the OAI-ORE protocol itself [22].

IV. Related work

Geospatial data have been traditionally disposed in collections. Raster data files covering the same area normally come as sets of files that form a collection, and in consequence metadata descriptions are also organized in nested collections [33] [34]. This imposes some degree of linkage since data are grouped according to a certain geographic criterion (proximity, overlay, etc.). The term collection here embraces resources and metadata together, no longer stored separately. Besides, as highlighted throughout this paper, we emphasize tremendous heterogeneity with regard to resources [35], say, geoprocessing services, multimedia resources, and raster data may be part of the same collection.

Regarding geoprocessing services, several research works deal directly with the OGC WPS specification [5] [7] [36]. In most cases, composing properly geospatial services remains still unsatisfactory due to the complexity inherent in some service specification (eg. WFS interface). Other alternatives to distributed geoprocessing computing are semantic-based [37], grid-enabled [38], and REST extensions for BPEL [39].

In the digital libraries domain, recent works have explored the connections between OAI-PMH and LOD communities [40]. OAI-ORE related development has been constantly increasing since its birth, not only at server but also at desktop level appearing different tools that allow users to create their own collections or compound objets. This is the case of LORE [41] a tool created for authoring and publishing compound objects or collections based on the OAI-ORE model for representing bibliographic relationships.

Recent works reveal an increasing interest in connecting geospatial resources of any type [42]. In the SDI community, Florczyk et al. (2010) are exploiting semantic linkages to services in SDIs [43]. The authors propose a linked ontology of administrative units for referencing the same geographic concept to the corresponding instances form multiple WFS services. Similar examples come from the Ordnance Survey, with the publication of the Administrative Geography of Great Britain, an initiative to publish administrative units as linked data sources [44]. Shade and Cox (2010) have recently pointed the similarities between LOD and SDI since geospatial data encoded in GML permit simple mappings to RDF [45]. Our approach built on the OAI-ORE data model goes in this line to provide better support for metadata of individual resources and entire collections. Each resource and its metadata form a logical unit no longer separated, which enables greatly the discovery, access, and linkage to resources and collections.

V. Conclusion and future work

This paper has presented an ongoing approach to conciliate geospatial services and data with external LOD datasets. The use and extension of OAI-ORE protocol to model collections of interlinked geoprocessing services allows users to regroup and restructure the spectrum of data and services over the Web, in order to build functional, cross-domain Web applications. What is novel here is the straightforward, direct method for describing and packaging resources and collections, compared with the family of business process languages (BPEL, etc.), which makes it easy to browse, compose, access and visualize collections of interlinked geospatial resources.

Applications consuming properly structured data are still missing [26]. In this sense, our future research efforts are centred on building suitable tools to support the creation of OAI-ORE collections and their visualization over virtual globe platforms. Other challenging tasks concern with connecting current SDI catalogue services and applications (geoportals, etc.) to LOD datasets in order to link the relatively small SDI community to others much bigger, so that geospatial researchers may tackle multidisciplinary projects that expand the boundaries of geospatial information.

Acknowledgment

This work has been partially supported by the CENIT "España Virtual" project funded by the CDTI in the program "Ingenio 2010" through *Centro Nacional de Información Geográfica* (CNIG), and the EuroGEOSS project (ref. 226487) under the EU FP7 programme.

References

[1] C. Granell, C. Abargues, L. Díaz, and J. Huerta, "Interlinking geoprocessing services," in *Proc. of the IEEE Intl Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2010)*, IEEE Press, 2010, pp. 99-104.

[2] M.F. Goodchild, "Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0," *Intl. J. of Spatial Data Infrastructures Research*, vol. 2, pp. 24-32, 2007.

[3] M. Craglia, "Next Generation Challenges," in *Proc. of the AGILE Intl. Conference on Geographic Information Science (AGILE 2009)*, Hannover, Germany, Jun. 2009.

[4] M. Craglia *et al.*, "Next-Generation Digital Earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science," *Intl. J. of Spatial Data Infrastructures Research,* vol. 3, pp. 146-167, 2008.

[5] T. Foerster, B. Schaeffer, J. Brauner, and S. Jirka, "Integrating OGC Web Processing Services into Geospatial Mass-Market Applications," in *Proc. IEEE Intl. Conference on Advanced Geographic Information Systems & Web Services (GEOWS 2009)*, IEEE Press, 2009, pp. 98-103.

[6] T. Foerster, B. Schaeffer, S. Jirka, and J. Brauner, "Integrating Web-based Sensor Information into Geospatial Mass-market Applications through OGC Web Processing Services," *Intl. J. on Advances in Intelligent Systems*, vol. 2, no. 2&3, pp. 278-287, 2009.

[7] C. Granell, L. Díaz, and M. Gould, "Service-oriented applications for environmental models: reusable geospatial services," *Environmental Modelling & Software*, vol. 25, no. 2, pp. 182-198, 2010.

[8] J. Brauner, T. Foerster, B. Schaeffer, and B. Baranski, "Towards a Research Agenda for Geoprocessing Services," in *Proc. of the AGILE Intl. Conference on Geographic Information Science (AGILE 2009)*, Hannover, Germany, Jun. 2009.

[9] C. Bizer, R. Cyganiak, and T. Heath, "How to Publish Linked Data on the Web," Tech. Rep., 2007. [Online] Available: http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

[10] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *Intl. J. on Semantic Web and Information Systems*, vol. 5, no. 3, pp. 1-22, 2009.

[11] M.F. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, *Interoperating Geographic Information Systems*. Norwell, MA: Kluwer Academic Publishers, 1999.

[12] P. Zhao, G. Yu, and L. Di, "Geospatial Web Services," in B.N. Hilton (Ed.), *Emerging Spatial Information Systems and Applications*, pp. 1-35. Hershey, PA: IDEA Group, 2007.

[13] Open Geospatial Consortium Inc., "Web Map Service Implementation Specification 1.3.0," 2006. [Online] Available: http://www.opengeospatial.org/standards/wms

[14] Open Geospatial Consortium Inc.,"Web Feature Service Implementation Specification 1.1.0," 2005. [Online] Available: http://www.opengeospatial.org/standards/wfs

[15] Open Geospatial Consortium Inc.,"Web Coverage Service Implementation Specification 1.1.2," 2008. [Online] Available: http://www.opengeospatial.org/standards/wcs

[16] Open Geospatial Consortium Inc.,"OpenGIS Catalogue Service Implementation Specification 2.0.2," 2007. [Online] Available: http://www.opengeospatial.org/standards/specifications/catalog

[17] L. Bernard, I. Kanellopoulos, A. Annoni, and P. Smits,"The European Geoportal – one step towards the establishment of a European spatial data infrastructure," *Computers, Environment and Urban Systems*, vol. 29, no. 1, pp. 15-31, 2005.

[18] C. Granell, L. Díaz, and M. Gould, "Distributed Geospatial Processing Services," in M. Khosrow-Pour (Ed), *Encyclopedia of Information Science and Technology, Second Edition*, pp. 1186-1193. Hershey, PA: Information Science Reference, 2008.

[19] Open Geospatial Consortium Inc.,"OpenGIS Web Processing Service 1.0.0," 2007. [Online] Available: http://www.opengeospatial.org/standards/wps

[20] N. Alameh, "Chaining Geographic Information Web Services," *IEEE Internet Computing*, vol. 7, no. 5, pp. 22-29, 2003.

[21] K. Holtman and A. Mutz, "Transparent Content Negotiation in HTTP," Internet Engineering Task Force (IETF) Memo RFC 2295. [Online] Available: http://www.ietf.org/rfc/rfc2295.txt

[22] Open Archives Initiative, "Open Reuse and Exchange," 2008. [Online] Available: http://www.openarchives.org/ore/

[23] Open Archives Initiative, "ORE User Guide - Abstract Data Model," Oct 2008. [Online] Available: http://www.openarchives.org/ore/1.0/datamodel

[24] Open Archives Initiative, "Protocol for Metadata Harvesting," 2002. [Online] Available: http://www.openarchives.org/pmh/

[25] A. Maslov, A. Mikeal, S. Phillips, J. Leggett, and M. McFarland, "Adding OAI-ORE Support to Repository Platforms," in *Proc. of the 4th Intl. Conference on Open Repositories (OR'09)*, Atlanta, Georgia, 2009.

[26] M. Hausenblas, "Exploiting Linked Data to Build Web Applications," *IEEE Internet Computing*, vol. 13, no.4, pp. 68-73, 2009.

[27] M. Nottingham and R. Sayre, "The Atom Syndication Format," Internet Engineering Task Force (IETF) Memo RFC 4287. [Online] Available: http://www.ietf.org/rfc/rfc4287.txt

[28] C. Granell, M. Gould, R. Gronmo, and D. Skogan, "Improving Reuse of Web Service Compositions," in *Proc. of the 6th Intl. Conference on E-Commerce and Web Technologies (EC-Web 2005)*, LNCS, 2005 pp. 358-367.

[29] Open Geospatial Consortium Inc.,"OGC KML 2.2," 2008. [Online] Available: http://www.opengeospatial.org/standards/kml/

[30] Open Geospatial Consortium Inc.,"OpenGIS Geography Markup Language (GML) Encoding Standard 3.2.1," 2007. [Online] Available: http://www.opengeospatial.org/standards/gml

[31] M. Lutz, "Ontology-based descriptions for semantic discovery and composition of geoprocessing services," *GeoInformatica*, vol. 11, no. 1, pp. 1-36, 2007.

[32] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, L.A. Stein, and F.W. Olin, "OWL Web Ontology Language Reference," W3C Recommendation, 2004. [Online] Available: http://www.w3.org/TR/owl-ref/

[33] M.F. Goodchild and J. Zhou, "Finding geographic information: Collection-level metadata," *GeoInformatica*, vol. 7, no. 2, pp. 95-112, 2003.

[34] J. Nogueras-Iso, F.J. Zarazaga-Soria, and P.R. Muro-Medrano, *Geographic information metadata for spatial data infrastructures: resources, interoperability and information retrieval*. Berlin: Springer, 2005.

[35] Y. Raimond, C. Sutton, and M. Sandler, "Interlinking Music-Related Data on the Web," *IEEE MultiMedia*, vol. 16, no. 2, pp. 52-63, 2009.

[36] A. Friis-Christensen, R. Lucchi, M. Lut, and N. Ostlander, "Service chaining architectures for applications implementing distributed geographic information processing," *Intl. J. of Geographical Information Science*, vol. 23, no. 5, pp. 561-580, 2009.

[37] P .Yue, L. Di, W. Yang, G. Yu, and P. Zhao, "Semantics-based automatic composition of geospatial web service chains," *Computers & Geosciences*, vol. 33, no. 5, pp. 649-665, 2007.

[38] A. Chen, L. Di, Y. Wei, Y. Bai, and Y. Liu. "Use of grid computing for modeling virtual geospatial products," *Intl. J. of Geographical Information Science*, vol. 23, no. 5, pp. 581-604, 2009.

[39] C. Pautasso, "RESTful Web service composition with BPEL for REST," *Data & Knowledge Engineering*, vol. 68, no. 9, pp. 851-866, 2009.

[40] B. Haslhofer and B.Schandl, "Interweaving OAI-PMH Data Sources with the Linked Data Cloud," *Intl. J. Metadata, Semantics and Ontologies*, vol. 5, no.1, pp. 17-31, 2010.

[41] A. Gerber and J. Hunter, "A Compound Object Authoring and Publishing Tool for Literary Scholars based on the IFLA-FRBR," *Intl. J. of Digital Curation*, vol. 4, no. 2, pp. 28-42, 2009.

[42] M. Gahegan, J. Luo, S.D. Weaver, W. Pike, and T. Banchuen, "Connecting GEON: Making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure," *Computers & Geosciences*, vol. 35, no. 4, pp. 836-854, 2009.

[43] A. Florczyk, F.J. Lopez-Pellicer, R. Bjar, J. Nogueras-Iso, and F.J. Zarazaga-Soria, "Applying Semantic Linkage in the Geospatial Web," in *Proc. of the AGILE Intl. Conference on Geographic Information Science (AGILE 2010)*, Guimaraes, Portugal, May 2010.

[44] J. Goodwin, C. Dolbear, and G. Hart, "Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web," *Transactions in GIS*, vol. 12, no. s1, pp. 19-30, 2008.

[45] S. Schade and S. Cox, "Linked Data in SDI or How GML Is Not about Trees," in *Proc. of the AGILE Intl. Conference on Geographic Information Science (AGILE 2010)*, Guimaraes, Portugal, May 2010.