Semantic Augmentations for Geospatial Catalogue Service

Peng Yue[1, 2], Liping Di[2], Peisheng Zhao[2], Wenli Yang[2], Genong Yu[2], Yaxing Wei[2]
[1] Center for Spatial Information Science and Systems (CSISS), George Mason University,
6301 Ivy Lane, Suite 620, Greenbelt, MD 20770, USA
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University,
129 Luoyu Road, Wuhan, China, 430079

Abstract –**Catalogue service plays an important role in helping requestors to find the suitable geospatial data and services over the Web. The Open Geospatial Consortium has developed and recommended an ebRIM profile of Catalogue Services for the Web for implementing a catalogue service. Metadata for data and services registered in CSW is described by following the existing geographic metadata standards usually. The search functionality is limited to the direct match of keywords from metadata without fully utilizing the semantic information implicitly embedded in the metadata, such as hierarchical relationships among metadata entities. Web Ontology Language (OWL) provides a mechanism to enable the use of semantics. OWL-S uses OWL to describe the semantics for Web service. This paper explores the semantic representation of geospatial data and services to enable the semantic search in CSW based on the semantic relationship defined in OWL/OWL-S. Such semantics are organized in CSW through extending ebRIM elements. The paper also illustrates how such semantically-augmented CSW can facilitate service chaining and assist in dynamic discovery and/or derivation of geospatial information.**

Keywords: Geospatial, Catalogue Service for the Web (CSW), ebRIM, Ontology, OWL, OWL-S, Service Composition, Web Service

I. INTRODUCTION

Geospatial catalogue service provides capabilities for advertising and discovering shared data and services over the Web. Description Information (Metadata) for data and services are stored and organized in catalogue service to enable the search functionality. The Open Geospatial Consortium's (OGC) Catalogue Services for Web (CSW) is an industry consensus regarding an open, standard interface to online catalogs for geographic information and Web-accessible geoprocessing services. CSW specifies interfaces, HTTP protocol bindings, and a framework for defining application profiles required to publish and access digital catalogues of metadata for geographic data, services, and related resource information [1].

While CSW greatly facilitates the discovery of data and services, current discovery process is based on the static keyword match without the full exploration of underlying semantics, such as hierarchical relationships among metadata entities. Semantic augmentations to CSW, instead, can improve the discovery ability of data and services. Ontology has been used commonly as a knowledge representation mechanism for semantics in the computer science. An ontology is a formal, explicit specification of a shared conceptualization which provides a common vocabulary of an area and defines the meaning of the terms and the relations between them [2]. By formally conceptualizing metadata of data and services into ontologies, the semantics in metadata can be explicitly defined. Web Ontology Language[1] (OWL), as the standard Web ontology language recommended by W3C, provides the ability for explicit semantic representation. OWL Service Ontology [2] (OWL-S) as an OWL based ontology for Web services supports the capability description of services. These explicit specifications make the semantics of geospatial data and services machine understandable so that a flexible discovery of geospatial data and services based on semantics is possible. This paper explores the semantic representation of geospatial data and services to enable the semantic search in CSW based on the semantic relationship defined in OWL/OWL-S.

To make use of OWL/OWL-S in the discovery process of CSW, semantic relationship in OWL/OWL-S need to be stored in CSW. Two industry models exist for information registry: Universal Discovery Description and Integration (UDDI) model and Electronic Business Registry Information Model (ebRIM). UDDI deals only with services and its registry model is not flexible enough for data registration comparing to the ebRIM. The ebRIM profile for CSW implementation [3] introduces an ebRIM-based catalogue information model for publishing and discovering geospatial information. The ebRIM is a widely adapted information model that defines types of objects stored in a registry as well as the relationships among these object types. Geospatial data and services can be registered in this model following the geospatial metadata standards. However, current CSW specification doesn't take into account the registration of the semantic information in a CSW. Because ebRIM is a general and extensible registry information model which can be extended through its class, slot and association elements, we extend these elements to allow registering the semantic information of geospatial data and services.

Chaining and execution of geospatial Web services provides a flexible yet powerful way to derive high-level geospatial information from lower-level inputs through real-time integration of interoperable geospatial services and data. Based on the semantic description in CSW, this paper presents an automatic, "DataType"-driven service chaining process. The process produces an executable composite service to generate an on-demand geospatial product corresponding to user's requirements.

II. EBRIM-BASED INFORMATION MODEL IN CSW

[1] http://www.w3.org/2004/OWL/
[2] http://www.daml.org/services/owl-s/

Under the guideline of ebRIM profile for CSW, LAITS CSW [4] extended ebRIM with two international geographic standards: ISO 19115 [5] (including the draft part 2) and ISO 19119 [6]. Figure 1 shows a high-level view of the catalogue information model.

The ebRIM is extended with ISO 19115 and ISO 19119 in two ways. The first is deriving new metadata classes from existing ebRIM classes by importing new classes into the ebRIM class tree. For example, class CSWExtrinsicObject is derived from an existing class in the ebRIM – ExtrinsicObject – to represent all the metadata objects that describe those not intrinsic to the catalogue. Class Dataset is derived from CSWExtrinsicObject in order to describe geographic datasets. Many new attributes are added to the Dataset class based on ISO 19115 and its draft part 2.

The second way is using Slots to extend an existing class. Every class extended from class RegistryObject has the capability to add Slots. The ebRIM Service class can be used to describe geographic service but the available attributes in the class Service are not sufficient in describing geospatial Web services. Thus, new attributes derived from ISO 19119 are added to the Service class through Slots.

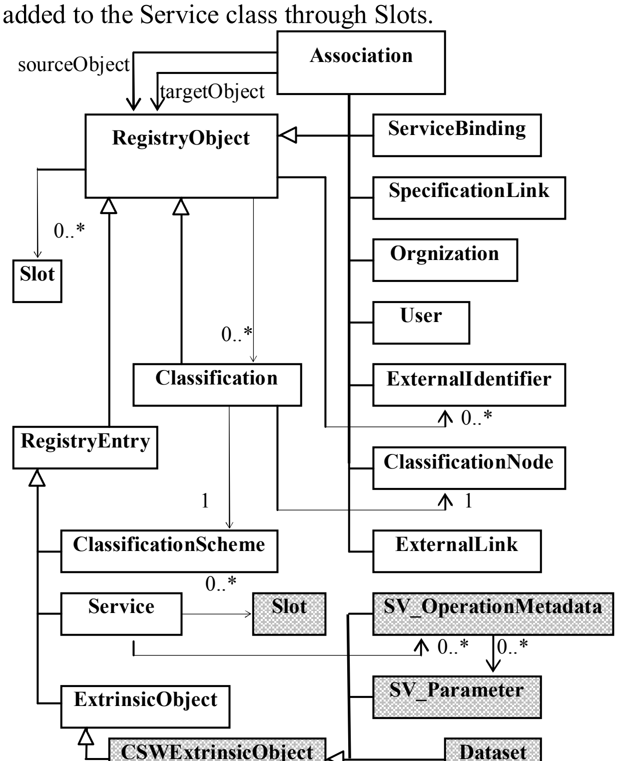

Figure 1. High-level view of catalogue information model

III. SEMANTICS FOR DATA AND SERVICES

Metadata for geospatial data includes much descriptive information of data such as identification, constraint, data quality, spatial/temporal representation, and content etc.[5] Corresponding semantic representation of the descriptive information is a large research topic. Our current research focuses on the semantic representation of data content, which involves the conceptualization of scientific theme keywords taxonomy using OWL. We refer to such OWL as "DataType" ontology. An example of such ontology derived from the conceptualization of the Global Change Master Directory (GCMD) science keywords is shown in Figure 2.

According to ISO 19119, three types of entities are identified for service metadata [6]: (1) Service Instance: service itself, hosted on a specific set of hardware and accessible over a network, (2) Service Metadata: description of service operation and address etc., and (3) Service Type: describes a service instance which is of a specific service type, e.g. Web Coverage Service (WCS), Web Coordinate Transformation Service (WCTS). GCMD provides a comprehensive hierarchical keyword list for services which we can conceptualize into "ServiceType" ontology. OWL-S as a service ontology is characterized by three modules: Service Profile, Process Model and Grounding. Service Profile describes the capabilities of Web services in terms of services' operation/functionality, input/output, pre/post-conditions. For example, a slope service has an input data type, the Digital Elevation Model (DEM) type, and may also have a precondition on data format (e.g., HDF-EOS format). At output, it generates a specific data type, i.e., the terrain slope type. OWL-S' support for automatic service discovery and composition is helpful for real-time production of requested data which otherwise do not exist. For example, if a slope data requested by a client does not exist, the service chaining capability may dynamically compose a chain by searching for a slope service and the input data needed to feed into this slope service through the semantic CSW. If the slope service is general enough to be applied to a broad spatial and temporal extents, the composed service chain can be viewed as a "Virtual data product", which can be cataloged in the CSW and be instantiated on demand. It has to be noted that currently OWL-S is still under development, thus we are not trying to incorporate all the metadata for service in OWL-S. Rather, we introduce OWL-S as an augmentation for service description to enable the production of "Virtual Data Product" through automatic service chaining. The "DataType" and "ServiceType" ontologies are used in the OWL-S for service description such as input/output "DataType", and service classification.

A service instance can be either tightly-coupled with a dataset instance, or un-associated with specific data instances, i.e. loosely-coupled [6]. Loosely-coupled services may have an association with data types instead of specific data instances. They can be described through Service Profile in OWL-S which advertises a certain type of services with specific input/output data types. In addition to "DataType" and "ServiceType" ontologies, we have also included the "Association" ontologies to describe the relationships between services and data. The introduction of association ontology can significantly speed up the reasoning process because it reduces the searching space through the association relationship expressed in the ontology. For example in figure 2, the "DataType" Terrain_Slope [1] is associated with

[1] Terrain_Slope is not conceptualized from GCMD keyword. It is defined through extending entity classes in the GCMD ontology.

"ServiceType" Image_Processing, then the searching process for services can start search within those services under the "Image_Processing" service type, which usually result quicker searching of the needed service. It serves as an optional search optimization strategy in the service discovery process for "Virtual Data Product". "DataType" and "ServiceType" ontologies act like a conceptual schema for semantic markups of dataset and service instances, hence we call it "Semantics Schema".

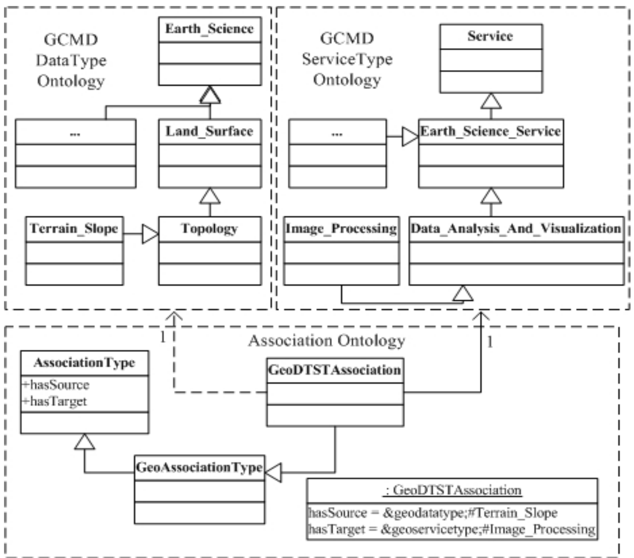

Figure 2. "DataType", "ServiceType" and "Association" ontologies

IV. SEMANTICS REGISTRATION IN CSW

In order to combine OWL/OWL-S and CSW, OWL/OWL-S has to be embedded into the ebRIM-based catalogue information model. A few recent studies have been reported regarding mapping OWL elements to ebRIM elements [7][8]. The basic idea is to use class, slot and association elements in ebRIM to record corresponding OWL classes, properties and related axioms such as subclassOf. In this study, we focus on the application and extension of ebRIM in the geospatial domain. This paper shows how to register the OWL/OWL-S for geospatial data and services in the ebRIM-based catalogue information model to support the semantic search functionality.

For data semantics, each dataset can be annotated in an extended ebRIM element described in the section II with the ontology entity class from OWL according to the data thematic information. For service semantics, the most straightforward way is to store the URI of OWL-S in an extended ebRIM element for service registry object. However, the capabilities advertised in OWL-S have to be extracted from URI by CSW to be compared with search conditions for services, which results in inconvenience. A more general approach is to build a mapping from the service profile of OWL-S to the underlying registry information model [9][10]. Table 1 shows current semantics registration in CSW and related explanations are listed.

(1) The "DataType" and "ServiceType" OWLs are registered in CSW as the ClassificationSchemes respectively.

(2) A new association type "assocGeoDTST" is defined with its sourceObject being ClassificationNode from the "DataType" ClassificationScheme and its targetObject being ClassificationNode from the "ServiceType" ClassificationScheme. Each association instance in the "Association" OWL is registered as an Association object under this association type.

(3) Each Service as an ebRIM RegistryObject is classified based on the "ServiceType" ClassificationScheme to specify its "ServiceType" through the associated ClassificationNode.

(4) Two new association type "assocServiceOutputDT" and "assocServiceInputDT" are defined with sourceObject being a Service object and targetObject being a ClassificationNode from the "DataType" ClassificationScheme. Each service instance is associated with its input and output "DataType"s through the Association objects under these two association types. Pre- and post-conditions are stored in the extended service slots.

(5) The Service class is extended with a slot to specify the URI of an OWL-S file for each service instance. When a match is found based on the service capabilities, the interaction with the service can be initiated through the specifications in the Process Model and Grounding parts of OWL-S. We keep these interaction details in the OWL-S file.

(6)Each Dataset as an ebRIM CSWExtrinsicObject is classified based on the "DataType" ClassificationScheme to specify its "DataType" through the associated ClassificationNode.

Table 1. Semantics registration in CSW

Geospatial Semantics	CSW
Semantics Schema	
"DataType" OWL	ClassificationScheme
"ServiceType" OWL	ClassificationScheme
"Association" OWL	Association
Service Semantics(OWL-S)	
"ServiceType"	ClassificationNode
Input "DataType"	Association
Output "DataType"	Association
Pre-Condition	Service Slot
Post-Condition	Service Slot
OWL-S URI	Service Slot
Data Semantics	
"DataType"	ClassificationNode

V. SEMANTIC SEARCH FUNCTIONALITY

Referring to the research in UDDI from the semantic Web, three options are available now for the implementation of semantic search functionality: (1) the component is created outside of CSW as a complement without any change to the CSW schemas [9]; (2) the component is embedded into CSW with some changes to the CSW schemas to support the semantically-augmented query [10]; and (3) the component is wrapped as an individual external matching services registered into CSW. In this option, CSW relay the matching task to the external matching services to enable the different types of matching [11]. We adopt the first option in our current study to minimize the impact on the existing CSW specification and implementation. Figure 3 provides a simplified view of the interaction between the matching components and CSW in order to respond to user's request.

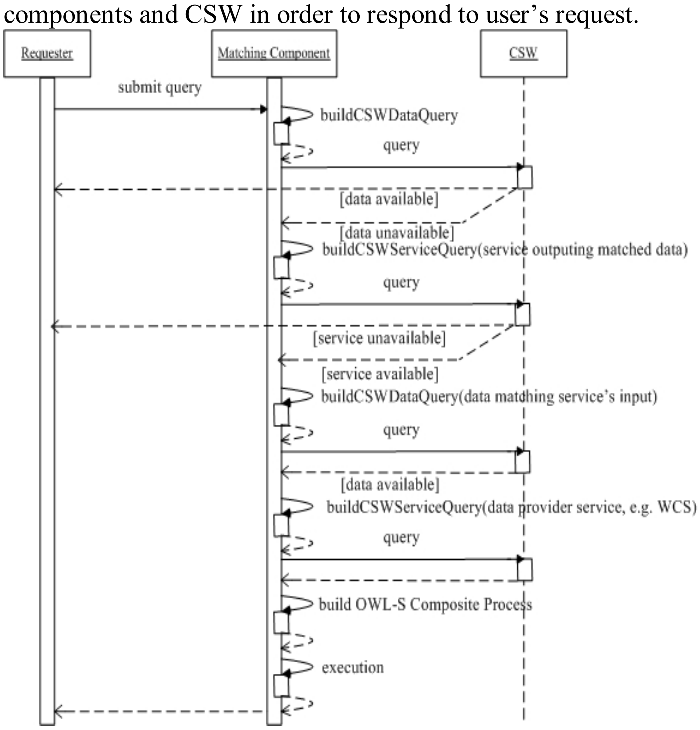

Figure 3. UML sequence diagram illustrating the searching process

Currently, semantic match is performed based mainly on the hierarchical relationships of ontology which includes EXACT, SUBSUME, RELAXED. Let OntR denotes the requested concept and OntP denotes the provider concept, the three matching conditions can be expressed as the followings with the decreasing priority order:

EXACT: OntR=OntP or OntR equivalent to OntP

SUBSUME: OntP is a subclassOf OntR

RELAXED: OntR is a subclassOf OntP

Users can set one of these matching options through the interface of the matching component.

Three types of semantically-augmented search functions are achieved:

1. Dataset Search: it gets semantically-matched "DataType"s as the additional search condition in the standard CSW dataset query.

2. Service Search: it gets semantically-matched "ServiceType"s with the optional input/output "DataType"s as the additional search conditions in the standard CSW service query.

3. "DataType"-Driven Service Chaining: a simplified process is illustrated in an UML sequence diagrams in figure 3. The composition is based on a match, either between two services among which the output of the first service provides the input of the second service, or between data and services such that the data provides the input of the service.

VI. CONCLUSION AND FUTURE WORK

This paper demonstrates how semantic search capability can be included in the ebRIM-based OGC Catalogue Service for Web. The implemented semantic matching execution component supports the semantically-augmented search for data and services. Such a semantically-augmented CSW can support "DataType"-driven service chaining with which "Virtual Data Product" can be constructed and cataloged and on-demand instantiation of such product can be generated. Future work includes developing richer ontology to depict the semantics for data and services and improving the precision in the discovery of geospatial data and services.

ACKNOWLEDGEMENT

This work is supported by a grant from U.S. National Geospatial-Intelligence Agency NURI program (HM1582-04-1-2021, PI: Dr. Liping Di).

REFERENCE

[1] D. Nebert, ed., 2003. OpenGIS® Catalog Services Specification, Version 2.0, OGC 03-108, Open GIS Consortium Inc. 164 pp.

[2] T. Gruber, A Translation Approach to Portable Ontology Specification, Knowledge Acquisition 5(2), pp. 199-220, 1993.

[3] R. Martell, ed., 2004. OGC™ Catalogue Services–ebRIM(ISO/TS 15000-3) profile of CSW, Version 0.9.1, OGC 04-017r1, Open GIS Consortium Inc. 87 pp.

[4] Y. Wei, L. Di, B. Zhao, G. Liao, A. Chen, Y. Bai, Y. Liu, 2005. The Design and Implementation of a Grid-enabled Catalogue Service. 25th Anniversary IGARSS 2005, July 25-29, COEX, Seoul, Korea. pp. 4224-4227.

[5] ISO/TC 211, ISO19115:2003, Geographic Information — Metadata.

[6] ISO/TC 211, ISO19119:2005, Geographic Information – Services.

[7] A. Dogac, Y. Kabak, G.B. Laleci, 2004. Enriching ebXML Registries with OWL Ontologies for Efficient Service Discovery, Proceedings of the 14th International Workshop on Research Issues on Data Engineering: Web Services for E-Commerce and E-Government Applications (RIDE'04). 8 pp.

[8] W. Liu, K. He, W. Liu, 2005. Design and realization of ebXML registry classification model based on ontology: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05), 6 pp.

[9] N. Srinivasan, M. Paolucci, K. Sycara, 2004. Adding OWL S to UDDI, implementation and throughput. First International Workshop on Semantic Web Services and Web Process Composition, San Diego, USA 2004. 12 pp.

[10] R. Akkiraju, R. Goodwin, P. Doshi, S. Roeder, 2003. A Method for Semantically Enhancing the Service Discovery Capabilities of UDDI. In Proceedings of the Workshop on Information Integration on the Web, Eighteenth International Joint Conference on Artificial Intelligence (IJCAI), Mexico, pp. 87-92

[11] J. Colgrave, R. Akkiraju, R. Goodwin, 2004. External Matching in UDDI. IEEE International Conference on Web Services, San Diego, USA, 2004. 8 pp.