Geoinformatica (2011) 15:273–303
DOI 10.1007/s10707-009-0096-1

Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure

Peng Yue · Jianya Gong · Liping Di · Lianlian He · Yaxing Wei

Received: 28 March 2009 / Revised: 18 August 2009
Accepted: 14 September 2009 / Published online: 9 October 2009

Abstract A geospatial catalogue service provides a network-based meta-information repository and interface for advertising and discovering shared geospatial data and services. Descriptive information (i.e., metadata) for geospatial data and services is structured and organized in catalogue services. The approaches currently available for searching and using that information are often inadequate. Semantic Web technologies show promise for better discovery methods by exploiting the underlying semantics. Such development needs special attention from the Cyberinfrastructure perspective, so that the traditional focus on discovery of and access to geospatial data can be expanded to support the increased demand for processing of geospatial information and discovery of knowledge. Semantic descriptions for geospatial data, services, and geoprocessing service chains are structured, organized, and registered through extending elements in the ebXML Registry Information Model (ebRIM) of a geospatial catalogue service, which follows the interface specifications of the Open Geospatial Consortium (OGC) Catalogue Services for the Web (CSW). The process models for geoprocessing service chains, as a type of geospatial knowledge, are captured, registered, and discoverable. Semantics-enhanced discovery for geospatial data, services/service chains, and process models is described. Semantic search middleware that can support virtual data product materialization is developed for the geospatial catalogue

P. Yue (✉) · J. Gong
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan, China 430079
e-mail: geopyue@gmail.com

L. Di
Center for Spatial Information Science and Systems (CSISS), George Mason University, 6301 Ivy Lane, Suite 620, Greenbelt, MD 20770, USA

L. He
Department of Mathematics, Hubei University of Education, Nanhuan Road 1, Wuhan, Hubei, China 430205

Y. Wei
Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6407, USA

service. The creation of such a semantics-enhanced geospatial catalogue service is important in meeting the demands for geospatial information discovery and analysis in Cyberinfrastructure.

Keywords CSW · ebRIM · Semantic · Cyberinfrastructure · Service chain · Geoprocessing workflow

1 Introduction

With the development of increasingly powerful sensor and platform technologies, geospatial users are experiencing a data-rich yet analysis-poor period. For example, NASA's Earth Observing System (EOS) alone is generating about 3.5 terabytes of data each day, far more than Earth scientists can hope to analyze. As a result, much data may not even once have been analyzed after collection. A new information infrastructure, the so-called Cyberinfrastructure (in the United States) or e-Infrastructure (in Europe) [1], is being developed to support the next generation of geoscientific research. The traditional focus on discovery of and access to geospatial data is being expanded primarily to enable scientific research using the Cyberinfrastructure, with its heavy analysis and synthesis demands [2]. Typical activities involve distributed geoprocessing workflows that support information processing and knowledge discovery from vast, heterogeneous data sets. The Web service technologies that allow scientists to set up this infrastructure for collaborative sharing of such distributed resources as geospatial data, processing modules, and process models are the technologies most widely used to support the Cyberinfrastructure. These technologies provide promise that users will be able to dynamically and collaboratively develop interoperable, Web-executable geospatial service modules and models, and apply them on-line to any part of the peta-byte archives to obtain customized information products rather than only raw data [3]. The Web service technologies follow the publish-find-bind paradigm in Service-Oriented Architecture (SOA) and have service discovery, description, and binding layers [4]. In the geospatial domain, a geospatial catalogue service provides a network-based meta-information repository and interface for advertising and discovering shared geospatial data and services[1]. The most widely used interface specification for geospatial catalogue services is the Open Geospatial Consortium (OGC)'s Catalogue Services for the Web (CSW).

Dynamically and collaboratively sharing and using resources is the concern of the Semantic Web community [5]. Semantic Web technologies, which give machine-processable meanings to the documents, allow the semantics of data and services to be used by machines (reasoning) for more effective discovery, integration, and reuse of geospatial data and services. A set of core technologies recommended by the World Wide Web Consortium (W3C) already exists, among them, Resource Description Framework (RDF), Web Ontology Language (OWL), and SPARQL Protocol And RDF Query Language (SPARQL). The Semantic Web community works closely with the Artificial Intelligence (AI) community. Members of the Semantic Web community have applied ontology concepts developed in the AI community to Web Services and search for and manipulation of Web information. Thus, these technologies show considerable promise for better discovery methods by exploiting underlying semantics in the descriptions for geospatial data and services.

[1] Hereafter, service, if not specified, means Web service.

While geospatial catalogue services greatly facilitate the discovery of data and services, the current discovery process is based on a static keyword match. The lack of explicit semantics inhibits the dynamic selection of those data, services, and geoprocessing workflows needed for processing geospatial information and discovering knowledge in a data-rich distributed environment. This paper addresses how semantic descriptions for geospatial data, services, and geoprocessing service chains can be structured, organized, and registered in geospatial catalogue services to allow processing of geospatial information and discovery of knowledge in addition to the discovery of geospatial information. The concept of a *virtual data product* is introduced, where, in addition to the discovery of those physically archived data products, a semantics-enhanced geospatial catalogue is used to discover data products that do not really exist in any archive but are materialized by composing available geoprocessing services and data. The proposed approach uses the ebXML Registry Information Model (ebRIM) of a geospatial catalogue service to store semantic descriptions and search for geospatial data, services, and geoprocessing service chains based on them. The approach is compared with other efforts to add semantics to catalogue services. How this approach can address geospatial information processing and knowledge discovery demands in Cyberinfrastructure is discussed.

2 Geoprocessing workflows, geospatial process models and virtual data products

The geoprocessing algorithm provided by geospatial services may handle only a tiny part of the overall geoprocessing or may be a large aggregated processing. In both situations, the service should be well defined, have clear input and output requirements, and be independently executable. Such services can be chained to construct different *geoprocessing workflows* (or service chains)[2] for geospatial knowledge discovery. In a distributed data and information environment such as the World Wide Web, there are many independent data and service providers. A complex geoprocessing workflow may be scattered among multiple service providers. Therefore, standards for publishing, finding, binding, and execution of services are needed. By following the standards for interfaces, interoperability of different software systems is achieved. Web services developed by different organizations can then be combined to fulfill users' requests. Through the OGC Web Services (OWS) testbeds, OGC has been developing a series of interface specifications under the OGC Abstract Service Architecture, including the Web Feature Service (WFS), Web Map Service (WMS), Web Coverage Service (WCS), Sensor Observation Service (SOS), CSW, and the Web Processing Service (WPS).

Figure 1 illustrates the relation among geoprocessing workflows, geospatial process models, and virtual data products. From the *knowledge discovery* perspective, the geoprocessing workflow transforms raw data into knowledge-added data products. For example, a landslide susceptibility data product, generated from the workflow processing the Digital Elevation Model (DEM) data and Landsat Enhanced Thematic Mapper (ETM) imagery, is a product of knowledge discovery. It has a process model that contains the landslide susceptibility, slope, aspect, land cover and Normalized Difference Vegetation Index (NDVI) computation subprocesses. In each of the subprocesses, it has its own model, i.e., calculating the landslide susceptibility index from the terrain slope and aspect, land cover type, and vegetation growing condition (i.e. NDVI) data, deriving the terrain slope

[2] Thereafter, in the context of this paper we use the term "geoprocessing workflow" and "geoprocessing service chain" interchangeably.

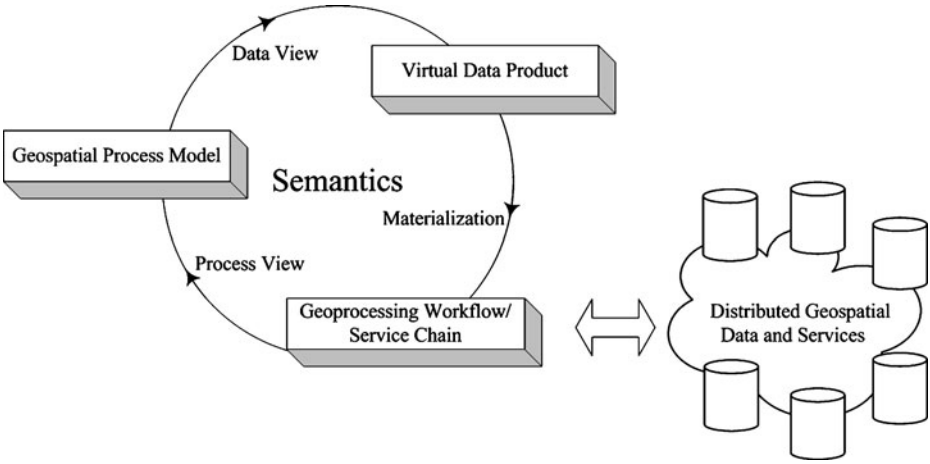

Fig. 1 Relation among geoprocessing workflows, geospatial process models, and virtual data products

and aspect from the DEM data, generating the land cover types using the image classification of the ETM imageries, and calculating ETM NDVI based on the Near-infrared (NIR) image (i.e. ETM Band 4) and red image (i.e. ETM Band 3). This Earth science application will be used throughout the paper to illustrate the proposed solution. The *process model* of a geoprocessing workflow contains knowledge from a specific application domain. In a service-oriented environment, the generation of geospatial process models means generating an abstract composite process model consisting of the control flow and data flow among process nodes. The data flow focuses on the data exchange among process nodes, while the control flow concerns the order in which process nodes are executed. A process node represents one type of many individual services that share the same functional behaviors such as functionality, input, and output. Using a process model, users can produce a required data product even though the product does not really exist in any archive; therefore, a process model produces a *virtual data product*, comparable to the physically archived data products. The virtual data product represents a geospatial data type that the process model can produce, not an instance (an individual dataset). It can be materialized on-demand as an executable geoprocessing workflow or a service chain when all required geoprocessing methods and inputs, often discovered through a geospatial catalogue service, are available. By defining domain concepts to represent the semantics of geospatial Web resources (whether data, Web services, or service chains), the linkage among geospatial data, services, and geoprocessing service chains can be used for more effective discovery, automation, integration, and reuse in various applications.

3 Critical issues for a semantics-enhanced geospatial catalogue service in the Cyberinfrastructure

Several critical issues must be explored when combining Semantic Web technologies and geospatial catalogue services to provide a semantics-enhanced geospatial catalogue service. Creation of a semantics-enhanced geospatial catalogue service in the Cyberinfrastructure requires

- Representation of semantics for distributed geospatial data, services, and geoprocessing workflows, using the ontology approach;

- Extension of the underlying catalogue information model of geospatial catalogue services to incorporate those semantics;
- Processing of catalogue queries using these semantics;
- On-demand delivery of geospatial information and knowledge through the integration of geospatial data, services, and geoprocessing workflows.

Resolution of these issues can be of great value in dealing with semantic heterogeneity, supporting geospatial data access, processing information, and discovering knowledge, thus contributing to the evolution of the Cyberinfrastructure.

3.1 Representation of semantics for geospatial data, services, and geoprocessing service chains

Ontology has been identified as a means to represent semantic knowledge in computer science. An ontology is "a formal, explicit specification of a conceptualization" [6] that provides a common vocabulary for an area and defines the meanings of terms and the relations between them. Ontologies have been used in the geospatial domain for information integration and semantic interoperability [7–9]. By mapping concepts in a geospatial Web resource (e.g. geospatial data, service, or geoprocessing service chain) to ontological concepts in the geospatial domain, the semantics of that geospatial resource can be explicitly defined.

OWL, recommended by W3C as a standard Web ontology language, is designed to enable the creation of ontologies and the instantiation of these ontologies in the description of Web resources [10]. OWL is an extension of the Resource Description Framework (RDF) [11], which defines a flexible approach to represent data, based on a graph model composed of triples. The foundation of the knowledge representation formalism for OWL is the description logic (DL) [12]. The basic elements of the description logic are *concepts*, *roles*, and *constants*. In the Web ontology context, they are commonly named *classes*, *properties*, and *individuals* respectively. Concepts group individuals into categories, roles stand for binary relations between those individuals and constants stand for individuals. The logical reasoning, called TBOX (Terminological Box) reasoning, supports determination of the subsumption, equivalence, and disjointness relations between concepts. For example, subsumption reasoning determines whether a concept is subsumed by another concept. In a geospatial ontology represented using OWL, if some "subClassOf" axioms are added to signify that "NDVI" is a sub-category of "Vegetation_Index" and "ETM_NDVI" is a sub-category of "NDVI", then DL reasoners can determine that "ETM_NDVI" is subsumed by "Vegetation_Index" using subsumption reasoning. The other type of reasoning, ABOX (Assertional Box) reasoning, is to determine whether a particular individual is an instance of a given concept description, or relations between individuals. For example, if a class "GeoTIFF" is defined to be a subclass of "MD_Format" with the only restriction that the inherited property "name_MD_Format" has a string value "application/GeoTIFF", DL reasoners can use ABOX reasoning to determine whether a particular individual of "MD_Format" is an instance of the class "GeoTIFF".

Semantic Web Services, the combination of the Semantic Web and Web Services, aim to provide mechanisms for organizing information and services so that the correct relationships between available data and services can be determined automatically, thus helping to build workflows for specific problems. The discovery of Web services is based on the capabilities that they provide such as inputs, outputs, preconditions and effects. W3C already recommends a standard for syntactic description of Web services: the Web Services

Description Language (WSDL). To address the semantics of Web services, the Semantic Web community has developed an OWL ontology for Web services known as OWL-S [13]. There are also other Semantic Web Service technologies available such as Web Service Modeling Ontology (WSMO) [14], Web Service Semantics (WSDL-S) [15], Semantic Annotations for WSDL (SAWSDL) [16], Semantic Web Services Framework (SWSF) [17]. WSMO and SWSF do not limit their knowledge representation to description logic. Thus, their definitions are not built upon OWL as OWL-S is. WSDL-S and SAWSDL aim to extend existing WSDL elements with semantic annotations; thus, they are not defining a complete ontology framework for Web services as OWL-S does. Most previous work uses OWL-S, and many tools are available. OWL-S can be selected as the starting point for the semantic description of geospatial Web services. OWL-S also provides a "Composite Process" ontology that contains the control and data flow among subprocesses. The control flow specifies the ordering and conditional execution of subprocesses, while the data flow focuses on data exchange among the subprocesses. Therefore, OWL-S can be used to describe the semantics of geoprocessing service chains. Section 4 will describe the usage of OWL/OWL-S for semantic descriptions of geospatial data, services, and geoprocessing service chains.

3.2 Extension of the underlying catalogue information model

OGC technology is the widely used choice for the standards-based interoperability and sharing technology of the Cyberinfrastructure for GIScience. The OGC CSW is an open industry consensus on a standard interface to online catalogs for geospatial data, services, and related resource information. Descriptive information (i.e., metadata) for geospatial information resources is structured and organized in catalogue services. The metadata can be queried and returned for evaluation, processing, and further binding or invocation of the cited resource. However, current standards mainly focus on syntactic interoperability and do not address semantic interoperability [18]. This work uses OGC standards to address the semantic interoperability of geospatial catalogue services.

Figure 2 shows the relations among the OGC catalogue services, CSW, and the ebRIM profile of CSW. The core elements in an OGC catalogue service are the information model, the query language, and the interface [19]. The *information model* describes the *information structures and semantics* of information resources. Therefore, the information model of catalogue services should address the content, syntax, and semantics of geospatial data, services, and geoprocessing service chains. The OGC catalogue specification is a general framework for catalogue service implementation. Application profiles can be derived from this base specification [19]. Interoperability among the different profiles requires the specification of a set of core metadata elements in the information model, in particular, the *core queryable properties* and *common returnable properties*. For example, the spatial extent is such a core metadata element. It is represented by a BoundingBox element in the core queryable properties and a coverage element (interpreted as the BoundingBox in the context of metadata for geospatial data and services) in the common returnable properties [19]. Queries based on these core queryable properties can be executed by any catalogue service, while the common returnable properties permit the use of metadata from any catalogue service. The *query language* assists in discovery of information resources in the catalogue. Different implementations of query languages, such as the OGC Filter Specification or Catalogue Interoperability Protocol (CIP) and GEO profiles of Z39.50 Type-1 queries, should support a minimum set of data types and query operations, the so-called OGC_Common Catalogue Query Language, to allow interoperability. For example,

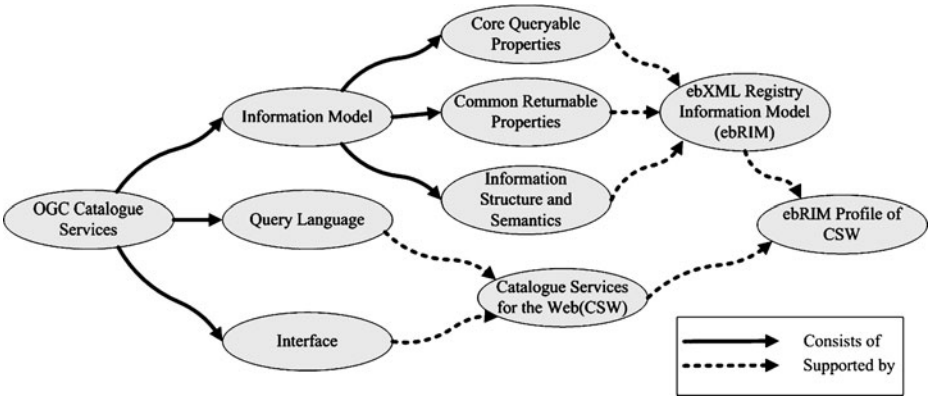

Fig. 2 OGC catalogue services and the ebRIM profile of CSW

the OGC_Common Catalogue Query Language defines spatial operators such as Intersects and Within that should be supported by all query language implementations to determine whether geometric arguments satisfy the claimed spatial relationship. The *interface* defines the functional behaviors of the catalogue service such as discovery and transactional operations. For example, it includes the getCapabilities operation, an operation supported by most OGC service specifications that allows clients to retrieve service metadata. Implementation of the interface in different distributed computing environments results in different protocol bindings, e.g., the CORBA protocol binding and HTTP protocol binding. CSW is a specification focusing on operations in the Web environment. It follows the HTTP protocol binding and can support XML encoding of the OGC Filter query language. The ebRIM standard has been defined by the Organization for the Advancement of Structured Information Standards (OASIS) and selected by OGC as the information model for specifying how catalogue content is structured and interrelated. Therefore, OGC proposes and recommends an ebRIM profile of CSW to join the CSW interfaces to ebRIM [20].

The ebRIM model is a general information model. It provides standard mechanisms to define and associate semantic information with registered information resources. Such mechanisms include using a cohesive set of extensibility points such as new kinds of associations, classifications, and additional slots (more details in Section 5). On the other hand, the Semantic Web is a separate effort. Semantics for geospatial data, services and geoprocessing service chains are represented using OWL/OWL-S. An important initiative for semantics-enhanced discovery of information resources based on ebRIM is to incorporate these explicitly defined semantics in OWL/OWL-S into ebRIM using these extensibility points. Various constructs in OWL are mapped to different ebRIM elements. Several efforts have already addressed this issue, although focusing only on the general information domain [21–24]. Section 5 will discuss how to make extensions to the ebRIM information model for geospatial catalogue services.

3.3 Semantics-based catalogue query formulation

The geospatial data, services and geoprocessing service chains must be discovered in the catalogue according to their semantics. In specifying query parameters, the requestor is best served by an ontology that defines these parameters in the context of the geospatial domain.

By having both the catalogue contents (as mentioned in Section 3.2) and query explicitly declare their semantics, the query results will be more relevant than discovery using static keyword matching.

There are basically two ways to use semantics to formulate catalogue queries. The first is to extend the catalogue interface to support a semantically-augmented query directly. One example is the work conducted by Akkiraju et al. [25], where they extended the Universal Discovery Description and Integration (UDDI) [26] inquiry API to incorporate RDF expressions in the query. UDDI provides an interface and information model for services registries. In contrast to ebRIM, UDDI deals only with services and its information model is not flexible enough for other information resources such as datasets. However, when making semantic enhancements to both types of registries, they do share some similarities such as scalable architecture design, query formulation, and semantic matching.

The other way to use semantics is to create middleware from the catalogue service, adding a semantic layer in front of it. Doing this allows reuse of legacy catalogue service, and there is no change to the interface of the catalogue service, while at the same time allowing a semantic matching between catalogue records and the query. Most existing work [27–30] follows this direction. In Section 5, semantic middleware is proposed to help dynamically generate semantics-based catalogue queries.

3.4 Supporting on-demand delivery of geospatial information and knowledge

Discovery of and access to the distributed geospatial data is highly useful, because it provides a global data repository that is as easily accessible to geospatial users as their local resources. But ultimately, it is the sharing and reusing of geospatial knowledge and automated generation of knowledge-added products that can be truly revolutionary, simply because it can support decision making directly and provide solutions instead of raw data. The semantics-enhanced catalogue service proposed in this paper should support the discovery of not only geospatial data and services, but also the process models, a kind of knowledge mentioned in Section 2. When archived data or services are not available, it can find existing process models or automatically generate new process models, link process models to geoprocessing workflows or service chains through data and services discovery, and automatically execute service chains to provide virtual data products that can meet the original demands. Therefore, the existing results of automatic service composition can be integrated into the proposed framework to support on-demand delivery of geospatial information and knowledge and provide virtual data products.

Semantic Web technologies have been widely used to support automatic service composition [39–41]. Specifically, they are usually combined with the AI technologies, especially AI planning methods. An important representation of planning problems related to the Web service field is using concepts of the state, goal and action from the classical planning domain. The world or a specified domain is modeled as a set of states that can be divided into initial states and goal states. Action is an operation that can change one state to another state. Thus, the assumption for Web service composition as a planning problem is that a Web service can be specified as an action with preconditions and effects. The preconditions are the states that must hold before the action can be executed, and the effects are the state changes when the action is executed [59]. First, a Web service is a software component that takes input data and produces output data. Thus, the preconditions and effects are the input and the output parameters of the service respectively. Second, the Web service might alter the states of the world after its execution. Then, the world states pre-required for the service execution are the preconditions, and the new states generated after

the execution are the effects [39]. The semantics for inputs, outputs, preconditions and effects (i.e. IOPE semantics) addressed in the Semantic Web Service technologies are widely used in most AI planning methods for automatic service composition [39–41], and then can be used in on-demand delivery of virtual data products.

4 Semantic descriptions for geospatial data, services, and geoprocessing service chains

The semantic descriptions for geospatial data, services, and geoprocessing service chains are based on the ontologies proposed by Yue et al. [31]. Our purpose is not to propose new ontologies for semantic descriptions of data, services and geoprocessing service chains. Rather we use the existing set of example ontologies and show how they can be exploited in a catalogue service. Here is a brief summary of the ontologies for geospatial data, services and geoprocessing service chains.

As shown in Fig. 3, geospatial DataType and ServiceType ontologies are defined for semantic descriptions of geospatial data, services, and service chains. Geospatial DataType ontology conceptualizes scientific meanings of distributed geospatial data, thus it can be

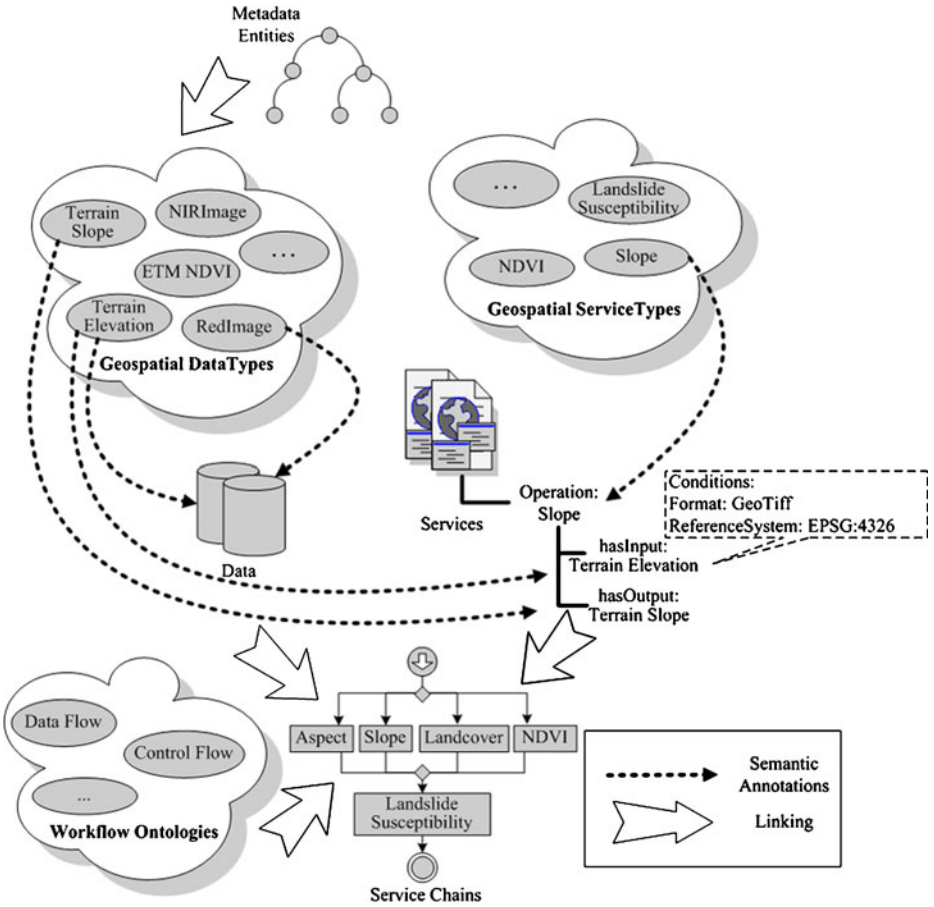

Fig. 3 Semantic descriptions for geospatial data, services and geoprocessing service chains

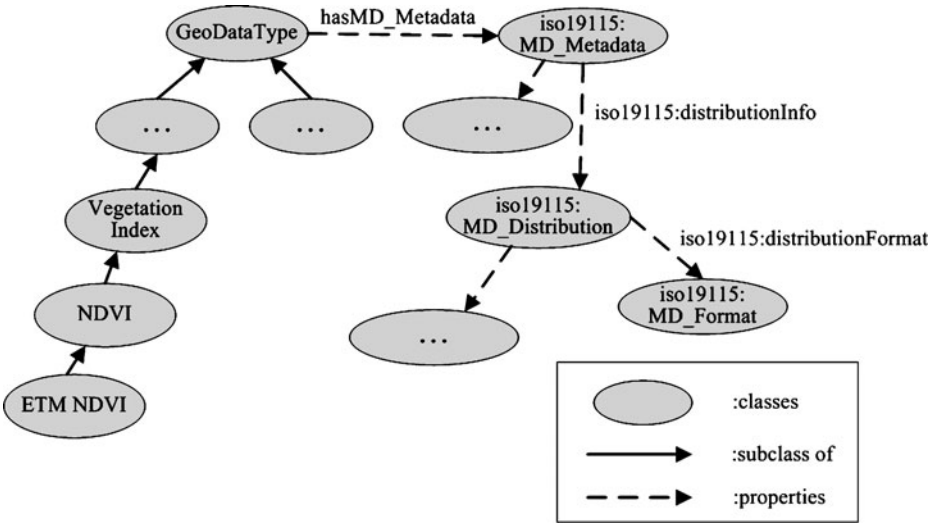

Fig. 4 An example geospatial DataType ontology

used to annotate the semantics of input and output data in a geospatial service operation. Furthermore, the DataType ontology can be enriched with metadata ontologies to allow more precise description of geospatial data, as shown in Fig. 4, and support cross-metadata-standards discovery [32] of geospatial data through additional semantic relations (e.g., "disjoint" and "equivalent") among terms in different metadata standards such as ISO 19115 [43] and the FGDC metadata standard. Geospatial ServiceType ontologies are defined according to the scientific problems that geospatial services focus on solving. Using geospatial DataTypes and ServiceTypes, we can represent the data, functional, and execution semantics of geospatial services [31]. Data semantics are the semantics of input and output data in a geospatial service operation, and thus are represented using geospatial DataTypes. The execution semantics of a geospatial service can be specified using the metadata statement in the preconditions and effects. For example, the preconditions for a slope computation service in Fig. 3 specify that the input terrain elevation data should be in the GeoTIFF data format with the EPSG:4326 geographic coordinate reference system. Geospatial ServiceTypes can be used to annotate the functionality of a geospatial service operation. Linking geospatial DataTypes, ServiceTypes, and workflow ontologies together, semantics for geospatial service chains are represented.

As mentioned in Section 3.1, Semantic Web technologies, in particular OWL and OWL-S, are used to represent semantics for geospatial data, services, and geoprocessing service chains. Geospatial DataType and ServiceType ontologies are expressed using OWL. OWL-S is used to express semantics for geospatial services and service chains. It consists of three main parts: service profile, service model (i.e. process), and service grounding. Table 1 shows a snippet of WSDL and OWL-S for the slope computation service. Geospatial DataType (e.g., Terrain_Elevation) and ServiceType (e.g., Slope) are linked into the OWL-S descriptions. The service grounding part of OWL-S provides information on how to bridge the syntactic and semantic worlds, e.g., grounding the input/output ontology concepts to the input/output message of WSDL using Extensible Stylesheet Language Transformations (XSLT) (Table 1). A process can be either atomic or composite. Both atomic and composite processes can be advertised through service profile ontology by their

Table 1 A snippet of WSDL and OWL-S for the slope computation service

```
<!--snippet of Slope WSDL -->
<message name="DEM2SlopeRequest"><part name="sourceURL"
   type="xsd:anyURI"/>…</message>
<message name="DEM2SlopeResponse"><part name="DEM2SlopeReturnURL"
   type="xsd:anyURI"/>…</message>
<portType name="SlopeCal">…
<operation name="DEM2Slope"><input message="DEM2SlopeRequest"/>
   <output message="DEM2SlopeResponse"/></operation></portType>

<!--snippet of OWL-S descriptions for Slope service-->
<!-- Service description -->
<service:Service rdf:ID="slope_service_01">
   <service:describedBy rdf:resource="#slope_process_01"/>
   <service:presents rdf:resource="#slope_profile_01"/>
   <service:supports rdf:resource="#slope_wsdlgrounding_01"/>
</service:Service>
<!-- Profile description -->
<profile:Profile rdf:ID=" slope_profile_01">
<profile:serviceClassification rdf:datatype="&xsd;#anyURI">&geoservicetype;#Slope
</profile:serviceClassification>…</profile:Profile>
<!-- Process Model description -->
<process:AtomicProcess rdf:ID="slope_process_01"> …</process:AtomicProcess>
<process:Input rdf:ID="slope_input_dem">
   <process:parameterType rdf:datatype="&xsd;#anyURI">
&geodatatype;#Terrain_Elevation</process:parameterType></process:Input>
<!-- Grounding description -->
<grounding:WsdlGrounding rdf:ID="slope_wsdlgrounding_01">
   <grounding:hasAtomicProcessGrounding rdf:resource="#
slope_wsdlatomicprocessgrounding_01"/></grounding:WsdlGrounding>

<!--snippet of service grounding-->
<grounding:wsdlInputMessage rdf:datatype="&xsd;#anyURI">
&slope_wsdl;#DEM2SlopeRequest</grounding:wsdlInputMessage>
<grounding:wsdlInput>
  <grounding:WsdlInputMessageMap rdf:ID="slope_wsdlinputmessagemap_dataurl">
   <grounding:owlsParameter rdf:resource="slope_input_dem"/>
   <grounding:wsdlMessagePart
rdf:datatype="&xsd;#anyURI">&slope_wsdl;#sourceURL</grounding:wsdlMessagePart>
   <grounding:xsltTransformationString><![CDATA[
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:iso19115="http://loki.cae.drexel.edu/~wbs/ontology/2004/09/iso-19115#"
xmlns:mediator="http://www.laits.gmu.edu/geo/ontology/domain/v3/mediator_v3.owl#"
xmlns:geodatatype="http://www.laits.gmu.edu/geo/ontology/domain/GeoDataType.owl#"
xmlns="http://slope.laits.gmu.edu">
          <xsl:template match="//geodatatype:Terrain_Elevation"> <xsl:value-of
select="mediator:hasMD_Metadata/iso19115:MD_Metadata/iso19115:distributionInfo/iso1911
5:MD_Distribution/iso19115:transferOptions/iso19115:MD_DigitalTransferOptions/iso19115:
onLine/iso19115:CI_OnlineResource/iso19115:linkage"/>
          </xsl:template></xsl:stylesheet>  ]]></grounding:xsltTransformationString>
</grounding:WsdlInputMessageMap><grounding:wsdlInput>
```

functionalities, inputs, outputs, preconditions, and effects. Atomic process ontology in OWL-S describes the behavior of an atomic service, while a composite process is a collection of subprocesses or atomic processes with control and data flow relationships. Therefore, the semantics for a geospatial service chain can be represented using composite process ontology. Table 2 illustrates semantic descriptions for the landslide susceptibility

Table 2 A snippet of OWL-S for a geoprocessing workflow

```
<!--snippet  of a composite process-->
<!-- Control Flow -->
<process:CompositeProcess …>
 <process:composedOf>
   <process:Sequence>
    <process:components>
     <process:ControlConstructList>
      <list:first>
       <process:Split-Join>
        <process:components>
         <process:ControlConstructBag>…
      <list:rest>
       <process:ControlConstructList>
        <list:first>
         <process:Perform rdf:nodeID="A0"/>
        </list:first>
        <list:rest rdf:resource="http://www.daml.org/services/owl-
s/1.1/generic/ObjectList.owl#nil"/>
       </process:ControlConstructList>
      </list:rest>
     </process:ControlConstructList>
    </process:components>
   </process:Sequence>
 </process:composedOf>
 <process:hasInput … />
 …
</process:CompositeProcess>
<!-- Data Flow -->
<process:Perform rdf:nodeID="A0">
  <process:process rdf:resource="&landslide_sus_4i;#landslide_sus_4i_process_01"/>
  <process:hasDataFrom>
   <process:InputBinding>
    <process:valueSource>
     <process:ValueOf>
      <process:fromProcess><process:Perform rdf:nodeID="A7"/></process:fromProcess>
      <process:theVar rdf:resource="&slope;#slope_output_slope"/>
     </process:ValueOf>
    </process:valueSource>
    <process:toParam rdf:resource="&landslide_sus_4i;#landslide_sus_4i_input_slope"/>
   </process:InputBinding>
  </process:hasDataFrom>
  <process:hasDataFrom>…
</process:Perform>
<process:Perform rdf:nodeID="…">…</process:Perform>
…
```

case using workflow ontologies in OWL-S. The control flow is represented by the control constructs such as Sequence and Split-Join. The data flow is specified through input/output bindings using a class such as ValueOf to state that the input to one subprocess should be the output of the previous one. For example, as shown in Table 2, the output (slope_output_slope) of the slope computation process is linked to the input (landslide_sus_4i_input_slope) of the landslide susceptibility atomic process.

5 Importing semantics into CSW

This section shows how the semantics for geospatial data, services, and geoprocessing service chains can be imported into CSW to enhance the discovery and processing of geospatial information.

5.1 The ebRIM information model for CSW

The ebRIM model specifies the metadata for information resources by using a set of classes and relationships among these classes. The UML style graph of Fig. 5 shows the relationships of the metadata classes defined by the model [33]. The core metadata class is the RegistryObject. Most other metadata classes in the information model are derived from this class. An instance of RegistryObject may have a set of zero or more Slot instances that serves as extensible attributes for this RegistryObject instance. An Association instance

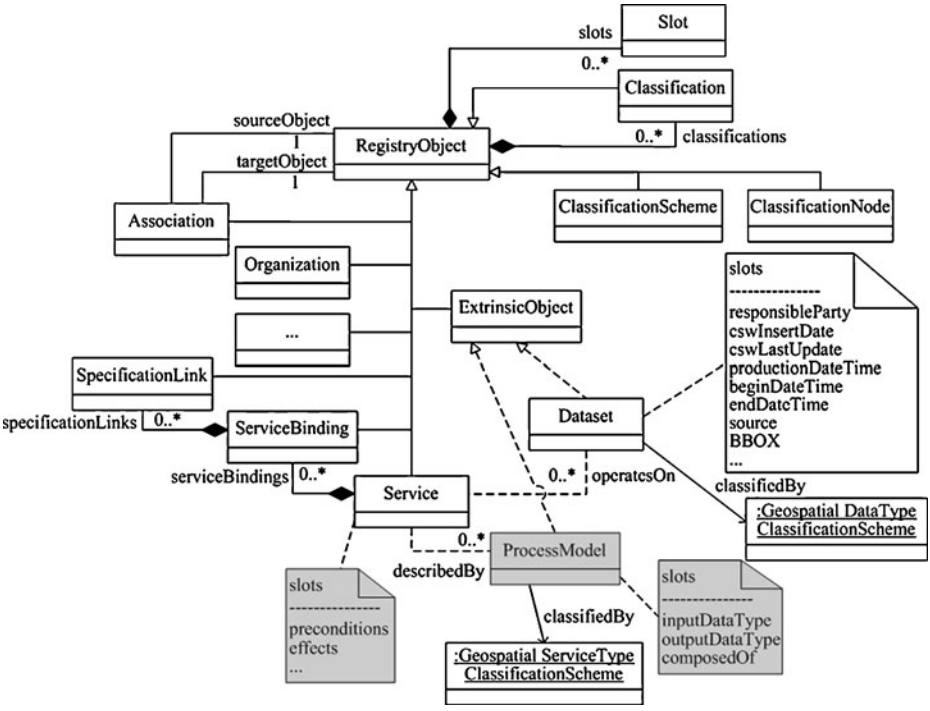

Fig. 5 Extension to the ebRIM information model

represents an association between a source RegistryObject and a target RegistryObject. Each association has an associationType attribute that identifies the type of that association.

The classification mechanism is a significant feature in the ebRIM information model. A Classification instance classifies a RegistryObject instance by referring to a node defined within a ClassificationScheme instance. A ClassificationScheme instance in the ebRIM model defines a tree structure made up of nodes that can be used to describe a taxonomy. The structure of a classification scheme may be defined internally to or externally of the registry, resulting in a distinction between internal and external classification schemes. The nodes in an internal classification scheme are instances of ClassificationNode. In an external classification scheme, the structure and values of the taxonomy elements are not known to the Registry. Classifications could be internal or external, depending on whether the classification scheme used is internal or external. The attributes in the Classification class allow for representation of both internal and external classifications [63]. An internal classification refers to a ClassificationNode in the internal ClassificationScheme, while an external classification refers to the node indirectly by specifying a representation of the node value unique within the external classification scheme.

The ebRIM model is a general standard model that can be adapted to meet specific requirements in the geospatial domain. The CSW-ebRIM profile [34] has provided guidance for registration of geospatial metadata by taking advantage of extensibility points offered by ebRIM. These extensibility points include new types of ExtrinsicObject, new kinds of associations, classifications, and additional slots. The dashed lines in Fig. 5 show extensions using these extensibility points. ExtrinsicObject provide metadata that describes submitted content whose type is not intrinsically known to the registry and therefore must be described by means of additional attributes. For example, metadata for geospatial data can be registered by creating a new type of ExtrinsicObject, i.e. Dataset (Fig. 5). New attributes such as spatial and temporal properties can be added to Dataset by defining additional slots. The ebRIM model has provided the Service class that supports the registration of service descriptions. A service chain as a whole can be conceived of as having a single-step execution that has inputs/outputs and performs a complex function. A WSDL can be also defined for a service chain. For example, the Web Services Business Process Execution Language (WSBPEL) [35], shortly known as BPEL, is an industry-wide standard that can be used for syntactic specification of service chains. An executable BPEL process can provide the process description for a service chain using activities, partners, and messages exchanged between these partners. A BPEL process works as a Web service and has a corresponding WSDL document. Therefore, descriptions of a service chain can also be registered in CSW using the Service class.

5.2 Extension of the ebRIM information model for registration of semantics

Extensions for registering semantics are created in the CSW-ebRIM profile. These extensions are designed to allow semantics-enhanced discovery and support on-demand delivery of geospatial data products. The following extensions shown as dark icons in Fig. 5 are made: 1) creating a new type of ExtrinsicObject, i.e. ProcessModel; 2) building new ClassificationScheme instances based on geospatial DataType and ServiceType ontologies; 3) adding slots to declare IOPE in the Service and ProcessModel classes.

5.2.1 Creating ProcessModel

As noted in Section 3.4, the semantics-enhanced catalogue service proposed in this paper supports the discovery of process models. Both atomic services and service chains have

process models that describe their behavior. A new association type DescribedBy, therefore, is defined with its sourceObject being a Service object and its targetObject being a ProcessModel object. The ebRIM model provides several standard classification schemes, such as ObjectType and AssociationType as a mechanism to provide extensible types. These classification schemes are called canonical classification schemes and can be extended by adding additional classification nodes. The ObjectType classification scheme defines the different types of RegistryObjects a registry may support, and therefore, the ProcessModel is defined as a classification node in this classification scheme, as shown in Table 3. The parent of the ProcessModel is a unique identifier referring to the parent classification node, namely ExtrinsicObject. The code of the ProcessModel contains a code that can be used in constructing the path. The path of the ProcessModel contains the canonical path from the root ClassificationScheme. The AssociationType classification scheme defines the types of associations between RegistryObjects. The association type DescribedBy is then defined as a classification node in the AssociationType classification scheme.

A Service instance can be either tightly coupled with a Dataset instance, or not associated with specific data instances, i.e. loosely coupled [18]. In the tightly coupled case, the service metadata describes both the service and the geographic dataset, the latter being associated to the service using the association type OperatesOn. Figure 6 shows an example of this association. Loosely coupled services may have an association with DataTypes instead of specific data instances. This type of association is conveyed through the process model for the service. As shown in the Fig. 7, the input/output data slots in the process model can address loosely-coupled associations. If the registered service is actually a

Table 3 The definition of ProcessModel in XML

```
<ClassificationScheme ...>
    <ClassificationNode ...>
        …
        <ClassificationNode xmlns="urn:oasis:names:tc:ebxml-regrep:xsd:rim:3.0"
xmlns:dsig="http://www.w3.org/2000/09/xmldsig#"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:oasis:names:tc:ebxml-regrep:xsd:rim:3.0
http://laits.gmu.edu:8099/csw/schema/rim-3.0.xsd" id="urn:uuid:7755e34b-c794-
4067-a409-7adb64bb6f7f" home="http://laits.gmu.edu:8099/csw/"
objectType="urn:uuid:555c406c-2850-4b34-b75f-fe936f670960" status="Approved"
parent="urn:uuid:6902675f-2f18-44b8-888b-c91db8b96b4d" code="ProcessModel"
path="/ObjectType/RegistryObject/ExtrinsicObject/ProcessModel">
            <Name>
                <LocalizedString xml:lang="en-US" charset="UTF-8"
value="ProcessModel"/>
            </Name>
            <Description>
                <LocalizedString xml:lang="en-US" charset="UTF-8"
value="process model for the service "/>
            </Description>
        </ClassificationNode>
    </ClassificationNode>
    …
</ClassificationScheme>
```

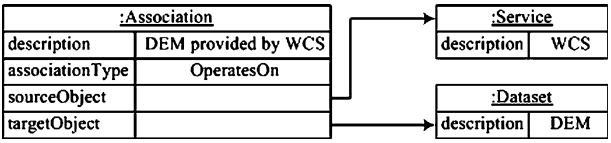

Fig. 6 Tightly coupled association between the service and data

composite service (i.e. service chain), the composedOf slot in the ProcessModel can link to a detailed composite process model such as the OWL-S composite process.

5.2.2 Building new ClassificationScheme instances

The ebXML Registry Profile for OWL proposed by the OASIS ebXML Registry Technical Committee [21] has provided a detailed guide on how to use ebRIM constructs to represent OWL constructs. An class in OWL must be mapped to a ClassificationNode in ebRIM. A classification scheme should be created for each ontology, and the classes belonging to this ontology should be represented as the classification nodes of this classification scheme. Therefore, two new ClassificationScheme instances as extensions are created, one for geospatial DataType ontology and the other one for geospatial ServiceType ontology.

Using these ClassificationSchemes, semantics can then be added by classifying geospatial data and services. Figure 8 shows that a dataset is classified according to the geospatial DataType classification scheme, using the associated classification node to specify its geospatial DataType. The lower part of Fig. 8 is an XML encoding example to illustrate this classification.

5.2.3 Adding slots for IOPE

IOPE semantics for geospatial services were illustrated in Section 4. The input and output semantics for geospatial services address the loosely-coupled type association between

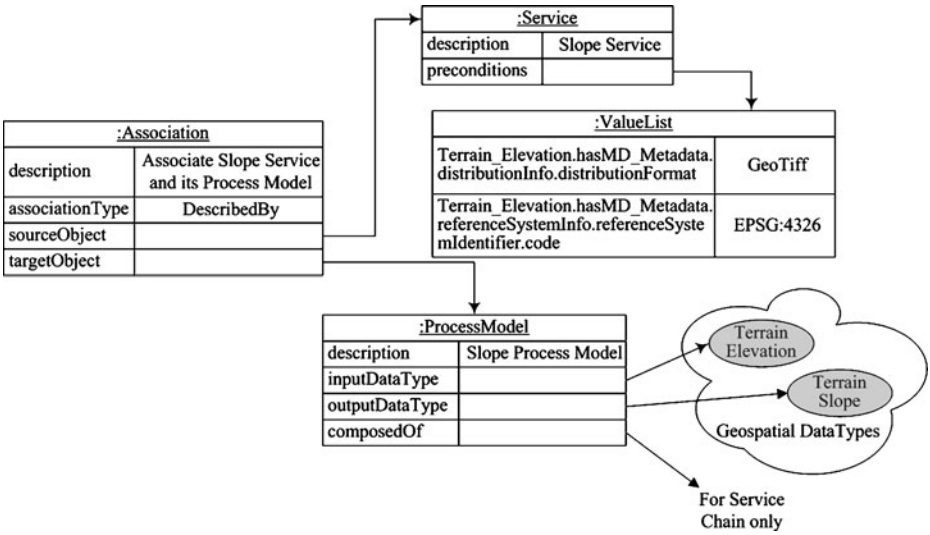

Fig. 7 An association between a service and its process model

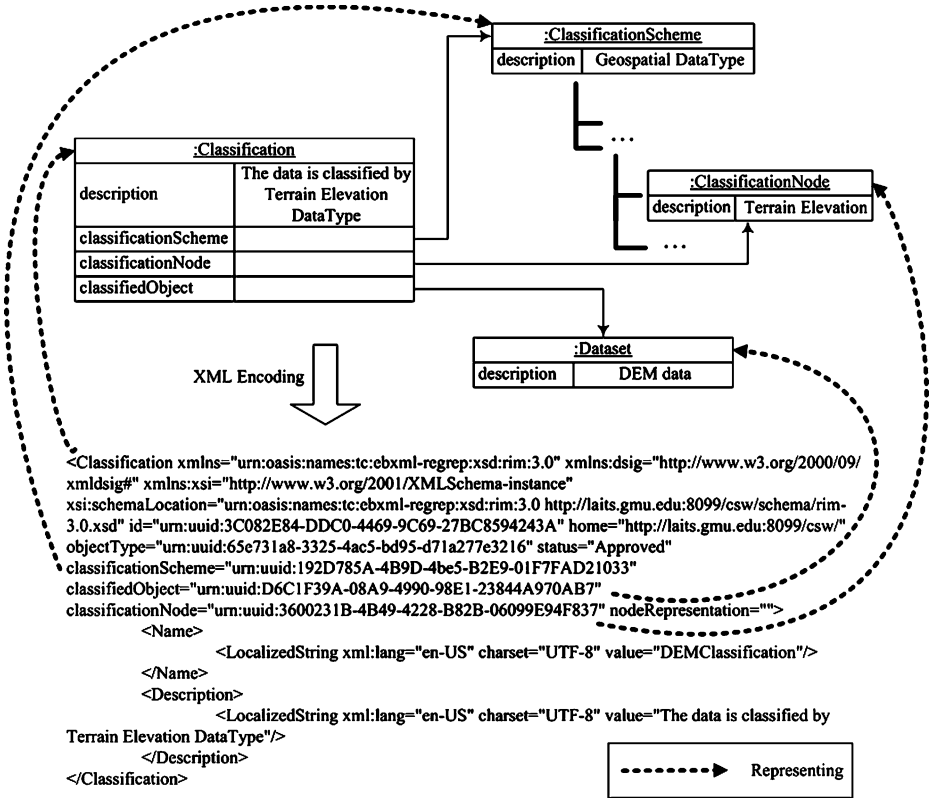

Fig. 8 Use of a geospatial DataType scheme for classification

services and data, and therefore are appropriate to be registered in the ProcessModel instances by adding inputDataType and outputDataType slots (Fig. 7). The values for these slots can be represented using ValueLists as shown in Table 4. Each value in the ValueList represents a unique identifier (e.g. URI) to the related geospatial DataType.

While input and output semantics address the loosely-coupled type association between services and data, the preconditions and effects for geospatial services proposed by Yue et al. [31] are concerned more with instance association between services and data. For example, many individual services may be available under the slope ServiceType; however, each service may have its own metadata requirements for the input data, such as a particular file format or spatial projection. This often happens in the geospatial domain and is due to the complex nature of geospatial data, which are highly multidisciplinary and heterogeneous. Such an association differs from the tightly-coupled association addressed in the association type OperatesOn, because no specific dataset is associated. We may call it a mixed-coupling case. The metadata constraints specified in the preconditions can be valuable when searching a more specific service under a ServiceType. As shown in Fig. 7, a ValueList can specify preconditions, where each value in the ValueList represents a c*ontextual path*, a term proposed by Bowers and Ludäscher [36], denoting a single concept, which may be in the context of one other concept through a series of properties. For example, "Terrain_Elevation.hasMD_Metadata.referenceSystemInfo.referenceSystemIdentifier.code" in Fig. 7 is such a path. Although its original purpose is to enable registration

Table 4 An example of inputDataType representation in XML

```
<Slot name="inputDataType" slotType="ProcessModel">
   <ValueList>
      <Value>http://www.laits.gmu.edu/geo/ontology/domain/GeoDataType.owl#
Terrain_Elevation</Value>
   </ValueList>
</Slot>
```

mappings and facilitate structural transformation of data, it does provide a way to identify a specific concept in a context, and thus can be used to identify a specific metadata element here. Figure 9 shows the mapping from a precondition represented using SPARQL to a contextual path.

5.3 Use of semantic extension for query formulation

The extended catalogue contents and DL-based reasoning are used to formulate queries. Those extended catalogue contents are queried through the standard CSW interface. Table 5 shows a geospatial data query using the standard GetRecords operation. The classification nodes and scheme for geospatial DataTypes are used as a search condition in the query. TBOX reasoning is used to derive additional concepts as the search conditions in the query. For example, those classification nodes with subclass-superclass relations determined by hierarchical relationships in the ontology (i.e., subsumption reasoning in TBOX reasoning) are added to the query conditions to allow a more effective discovery. If a user wants to find "Vegetation_Index" data, the query in Table 5 can be derived to search "ETM_NDVI" data that is semantically matched. As shown in Fig. 10, semantic middleware that can perform reasoning is created in front of the catalogue service, with no change to the legacy service interface. This semantic middleware is able to perform three types of discovery. The first is

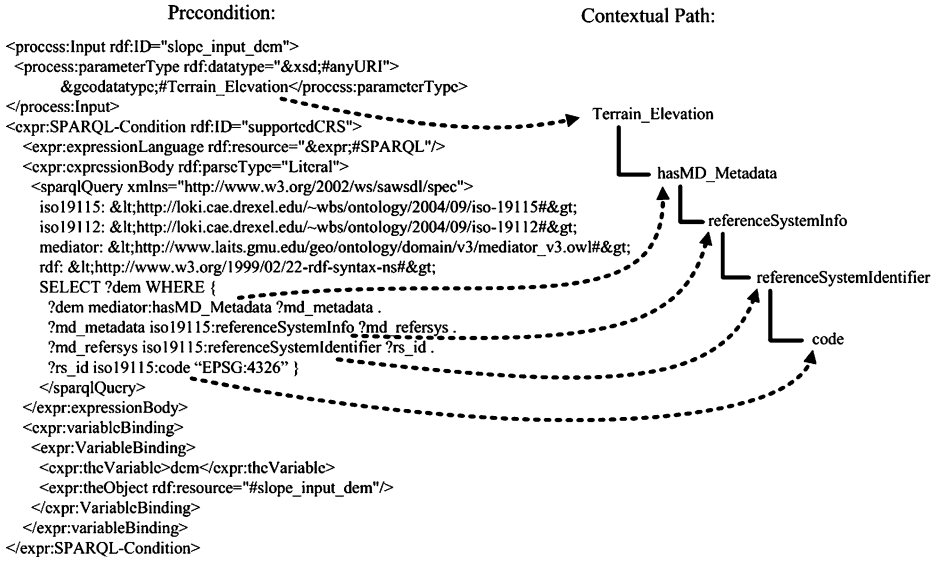

Fig. 9 Mapping from a SPARQL precondition to a contextual path

Table 5 A data query example, using geospatial DataType classification scheme

```
<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecords xmlns="http://www.opengis.net/cat/csw"
xmlns:csw="http://www.opengis.net/cat/csw" xmlns:ogc="http://www.opengis.net/ogc"
xmlns:gml="http://www.opengis.net/gml" version="2.0" outputFormat="text/xml"
charset="UTF-8" outputSchema="http://www.opengis.net/cat/csw" startPosition="1"
maxRecords="50">
  <csw:Query typeNames="Dataset Classification ClassificationScheme
ClassificationNode">
    <csw:ElementSetName>full</csw:ElementSetName><csw:ElementName>/Dataset/</c
sw:ElementName>
    <csw:Constraint version="1.0.0"><ogc:Filter><ogc:And>
       <!--temporal condition-->
        <ogc:PropertyIsGreaterThanOrEqualTo><ogc:PropertyName>/Dataset/beginDateTi
me</ogc:PropertyName>
          <ogc:Literal>2005-01-
10T00:00:00Z</ogc:Literal></ogc:PropertyIsGreaterThanOrEqualTo>
       <ogc:PropertyIsLessThanOrEqualTo><ogc:PropertyName>/
Dataset/endDateTime</ogc:PropertyName>
        <ogc:Literal>2005-01-
20T23:59:59Z</ogc:Literal></ogc:PropertyIsLessThanOrEqualTo>
       <!--spatial condition-->
       <ogc:BBOX><ogc:PropertyName>/ Dataset/BBOX</ogc:PropertyName>
        <gml:Box srsName="EPSG:4326">
          <gml:coordinates>-122.2167,37.7994 -
122.2167,37.7994</gml:coordinates></gml:Box></ogc:BBOX>
       <!—derived concept -->
       <ogc:PropertyIsEqualTo><ogc:PropertyName>/Dataset/@id</ogc:PropertyName>
        <ogc:PropertyName>/Classification/@classifiedObject</ogc:PropertyName></ogc:
PropertyIsEqualTo>
       <ogc:PropertyIsEqualTo><ogc:PropertyName>/Classification/@classificationSchem
e</ogc:PropertyName>
        <ogc:PropertyName>/ClassificationScheme/@id</ogc:PropertyName></ogc:Proper
tyIsEqualTo>
       <ogc:PropertyIsEqualTo>
        <ogc:PropertyName>/ClassificationScheme/Description/LocalizedString/@value</
ogc:PropertyName>
        <ogc:Literal>geospatial data type</ogc:Literal></ogc:PropertyIsEqualTo>
       <ogc:PropertyIsEqualTo><ogc:PropertyName>/Classification/@classificationNode<
/ogc:PropertyName>
        <ogc:PropertyName>/ClassificationNode/@id</ogc:PropertyName></ogc:Property
IsEqualTo>
       <ogc:PropertyIsEqualTo><ogc:PropertyName>/ClassificationNode/@code</ogc:Pro
pertyName>
        <ogc:Literal>ETM_NDVI</ogc:Literal></ogc:PropertyIsEqualTo>
     </ogc:And></ogc:Filter></csw:Constraint>
</csw:Query></csw:GetRecords>
```

geospatial data discovery using a classification scheme for geospatial DataTypes. The query in the Table 5 is such an example. The Rodriguez and Egenhofer [37] distance concept can be used as one option to control the enumeration of derived concepts. The distance is measured using the number of connected subclass-superclass arcs between two entity classes in the ontology, providing a reference value for assessing the similarity of entity classes.

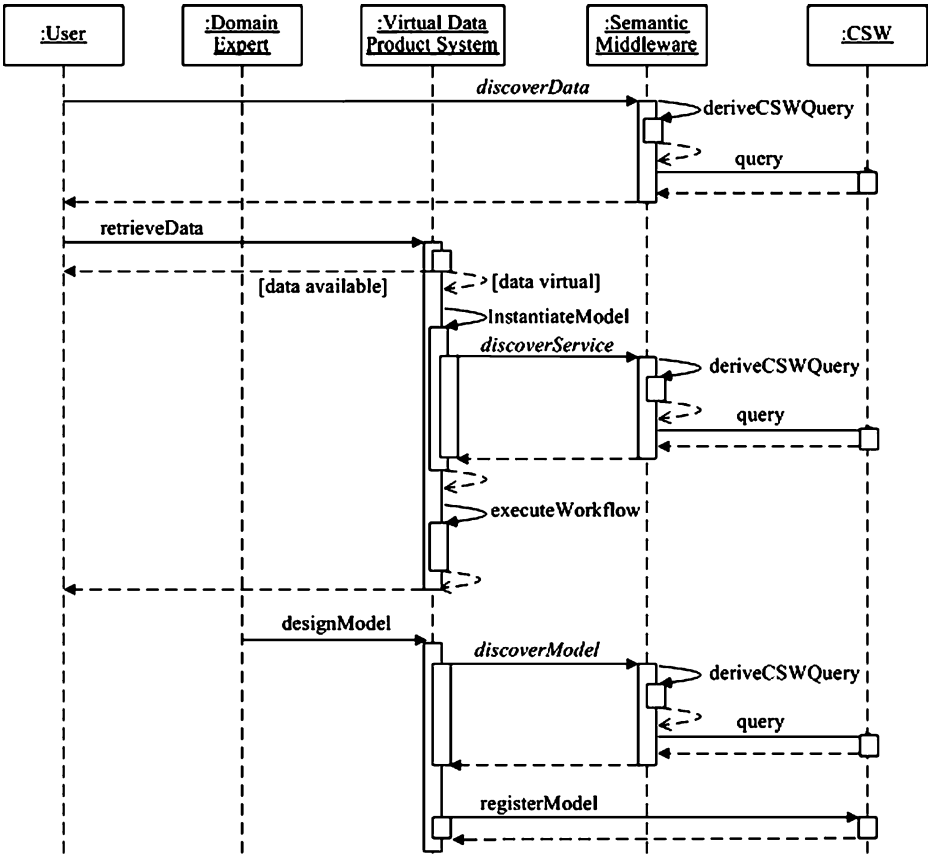

Fig. 10 UML sequence diagram illustrating the role of semantics-enhanced CSW in supporting virtual data production

The second type of discovery is service discovery. This kind of query includes discovering a service chain, since a service chain as a whole is a service. We have adopted the idea of a three-phase service discovery algorithm on UDDI by Sivashanmugam et al. [29], except that we introduce the concept of the process model and use a different information model here. The users' requirements for the data and functional semantics of services are constructed as process templates using ontological concepts. A process template is defined as a tuple (F, I, O), where F is the semantic concept addressing the function of the process, I is a finite set of input semantic concepts and O is a finite set of output semantic concepts. In the first phase, process models are discovered using a geospatial ServiceType classification scheme. This is similar to geospatial data query. In the second phase, the process models resulting from the first phase are ranked in order of semantic similarity [38] between the input and output concepts of the selected models and the input and output concepts of the template, respectively. In the third optional phase, users' requirements on the execution semantics of services (here metadata requirements) are constructed using contextual paths. The third phase then involves service discovery using process models resulting from the second phase and optional context paths. The first and second phases can be combined to support the third type of discovery, the process model

discovery. In addition, process model discovery can be flexible. For example, users might be interested in those models that can provide a certain output DataType. In this case, the output DataType serves as the only condition in the process model query.

It should be noted that the ebXML registry can use stored procedures to handle registered OWL semantics. For example, the path attribute in the ClassificationNode shown in Table 3 contains the canonical path leading from the parent nodes; therefore, it is feasible to derive semantically related geospatial DataTypes in the geospatial DataType classification scheme from this representation (e.g. path = "/GeoDataType/…/Vegetation_Index/NDVI/ETM_NDVI") by using string comparison functions in the relational database. Predefined queries can be defined to invoke such stored procedures. However, use of stored procedures can only achieve limited reasoning functionality since the semantics in OWL can be fully explored only in its own syntax-aware reasoners due to its intrinsic logic nature [22].

5.4 On-demand delivery of geospatial information and knowledge

A virtual data product system that is capable of deliver geospatial information and knowledge on-demand can combine the discovery of geospatial data, services, and process models into three consecutive phases for automatic service composition: process modeling, process model instantiation, and workflow execution [31, 39–41]. In the process-modeling phase, the knowledge of a domain expert is captured through process models. The model design process can be manual or automatic. In manual model design, users can find existing process models, link different geospatial DataTypes and ServiceTypes together to create new process models, compose new process models from existing process models, or check whether existing process models are decomposable or not. To automate this design process, the ontology reasoning introduced in Section 5.3 and AI planning methods can be used to automatically generate new process models. In both manual and automatic model design, discovery of existing process models must be involved. When the process model design has been completed and evaluated through process model instantiation and workflow execution, the model can be registered in the catalogue for future use. In the process model instantiation phase, the process model can be bound to a concrete geoprocessing workflow or executable service chain through data and services discovery. A workflow execution engine, then, can use a service chain to generate on-demand data products. Through these three phases, a virtual data product is then materialized to a data instance. The role of a semantics-enhanced geospatial catalogue service in supporting virtual data production is illustrated in the UML sequence diagram of Fig. 10.

6 Prototype implementation and result analysis

6.1 Implementation

Using the guidelines of the ebRIM profile for CSW, the CSW [42] implementation[3], developed and maintained by Laboratory for Advanced Information Technology and Standards (LAITS) from George Mason University, has extended ebRIM using international geographic standards: ISO 19115 Geographic Information - Metadata [43] (including part 2: Extensions for imagery and gridded data) and ISO 19119 Geographic Information—

[3] Online services are available at http://geobrain.laits.gmu.edu/

Services [18]. The ebRIM is extended with ISO 19115 and ISO 19119 in two ways. The first is by importing new classes into the ebRIM class tree, deriving new metadata classes from existing ebRIM classes. The new Dataset class is used to describe geographic datasets. Many new attributes are added to the Dataset class based on ISO 19115 and its part 2. The second way to extend ebRIM is to use Slots to extend an existing class. The Service class included in ebRIM can be used to describe geographic services, but the available attributes in the class Service are not sufficient to describe geospatial Web services. New attributes derived from ISO 19119 are added to the Service class through Slots.

Jena[4] and the OWL-S API[5] are used to construct the semantic middleware. Jena is a Java framework for building Semantic Web applications. It provides a programmatic environment for RDF/RDFS and OWL/SPARQL and includes a rule-based inference engine. By using Jena, one can parse, create and search the concepts in semantic models based on RDF technique. Jena Transitive and the OWL-Micro Reasoner[6] have been selected for reasoning. The Jena Transitive Reasoner is preferred because of its efficient TBOX reasoning. The OWL-S API provides a Java API for programmatic access for reading, executing, and writing OWL-S service descriptions. The API provides an ExecutionEngine that can invoke AtomicProcesses that have WSDL groundings and CompositeProcesses that use control constructs such as Sequence, and Split-Join. It has been extended in GMU LAITS to support the HTTP GET and POST invocations in addition to the SOAP invocation it has. The most advanced version of OWL-S that OWL-S API supports currently is version 1.1. It is also extended in this work to support some new features in the pre-release version of 1.2 including the support of the SPARQL precondition. OWLSManager, a system for the management of OWL-S files that can deploy and undeploy OWL-S files into the knowledge base, is developed [62]. The semantic middleware for CSW is integrated with the OWLSManager to support automatic geospatial service composition. Figure 11 shows the interface of OWLSManager for generating a virtual data product through automatic service composition, with the request of a virtual data product using XML and a semantics-enhanced CSW at the backend.

Yue et al. [27] have presented an architecture and implementation for a semantics-enhanced catalogue service based on GMU LAITS CSW. It extends the ebRIM information model to support registration of semantics for geospatial data and services and provides middleware to support the semantics-augmented discovery of geospatial data and services. However, support for geoprocessing service chaining is limited. The process models for geospatial services and chains are not captured and registered in the catalogue, thus the implementation cannot handle this kind of geospatial knowledge and support advanced automatic geospatial information processing and knowledge discovery. To provide support to a virtual data product system such as the one demonstrated by Zhao et al. [44], the original implementation is extended by adding process models and adjusting the discovery process correspondingly. Thus, a process model can be reused or generated as a new kind of geospatial knowledge to support geospatial information processing and knowledge discovery in Cyberinfrastructure.

6.2 Result analysis

To run the landslide susceptibility case in the OWLSManager, we have implemented all related Web services and created OWL-S descriptions for these online

[4] http://jena.sourceforge.net
[5] http://www.mindswap.org/2004/owl-s/api/
[6] http://jena.sourceforge.net/inference/index.html

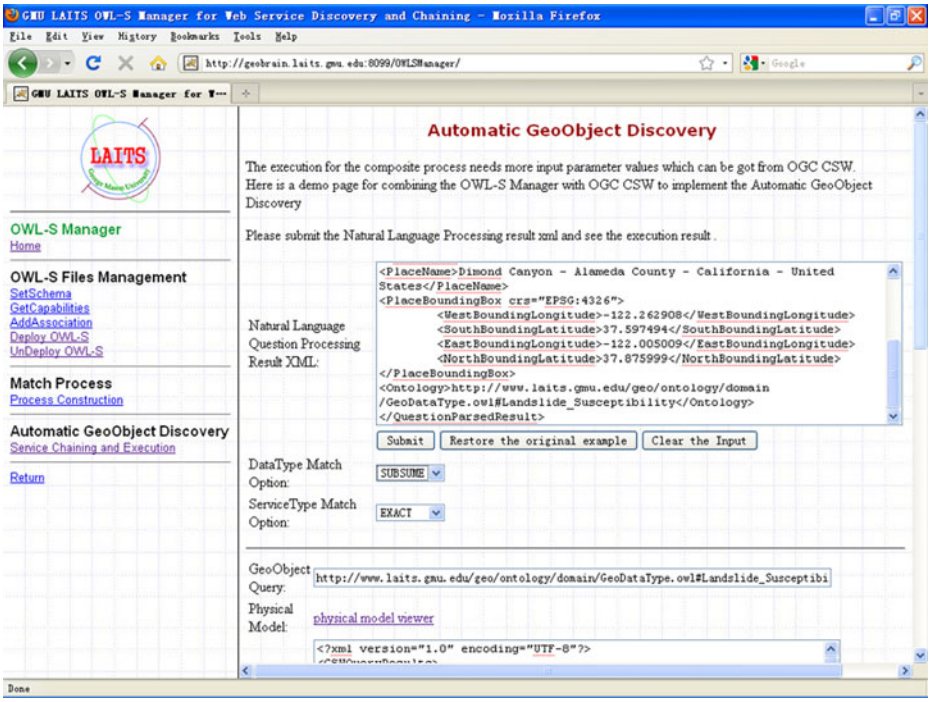

Fig. 11 OWLSManager interface, showing the request of a virtual data product

services[7]. Two atomic process models can provide a landslide susceptibility data product. The first atomic process model uses four types of data (slope, aspect, land cover, and NDVI) to calculate landslide susceptibility, and the second one only uses only slope and aspect data to calculate landslide susceptibility. Therefore, one atomic process model for landslide susceptibility can be combined with other discovered atomic processes. These processes can provide input data for that landslide susceptibility atomic process to create different composite process models. For example, an ETM NDVI calculation process that provides ETM NDVI data can be linked to the first landslide susceptibility atomic process mentioned based on subsumption reasoning.

The applicability of the semantics-enhanced CSW is demonstrated through its support to automatic chaining of multiple Web services to derive the landslide susceptibility index of the certain area (Diamond Canyon, California) on a certain day. A virtual data product request is represented using a geospatial DataType along with spatial and temporal conditions (Fig. 11). An atomic process model for landslide susceptibility (with either two or four inputs) is selected. Its input DataTypes are provided by slope and aspect (and land cover and NDVI) processes, whose input data are available and directly served through WCS. The service chain in this use case can be automatically and dynamically generated whenever the CSW service is available and queries can be augmented with the semantic extensions in the ebRIM information model. The composite process can also be registered

[7] All Web services, ontologies and related resources for this case are available at http://www.laits.gmu.edu/geo/nga/landslidecase.html

in the CSW as a virtual data product so that the composition process need not be repeated when a new request for the same data product is submitted. In addition to landslide susceptibility data, slope data, slope aspect data, landcover data or ETM NDVI data can also be created on demand. In this system, the service user is assisted by the ontologies from the knowledge base when selecting services and does not need to deal with syntactical service descriptions and WSDL message element mappings among possibly chainable services. That can help the domain expert focus more on the domain knowledge contribution instead of delving into the technical details. The prototype system demonstrates that both individual geoprocessing services and valid process models can be shared. The system is thus a self-evolving system whose capability will increase significantly as more individual services and modeling processes are inserted and/or developed, thus it contributes to the evolution of the Cyberinfrastructure.

7 Related work and discussion

It is envisioned in the U.S. National Science Foundation's (NSF) report that Cyberinfrastructure will be a comprehensive information infrastructure that integrates computing hardware and systems, data and information resources, networks, digitally enabled-sensors, online instruments and observatories, virtual organizations, and experimental facilities, along with an interoperable suite of software and middleware tools and services [60]. Grid technology, a distributed computing technology that involves the coordination and sharing of computing, application, data, storage, and network resources across dynamic and geographically dispersed organizations [61], plays an important role towards the Cyberinfrastructure currently. However, present-day versions of middleware for Grid and high-performance computing provide only a set of low-level middleware and a small part of the functionality required for Cyberinfrastructure. There are substantial research challenges to develop high-level intelligent middleware services and domain-specific services for problem-solving and scientific discovery in the Cyberinfrastructure [1]. Our work tries to provide intelligent geospatial services for geospatial information discovery and processing through integrating the Semantic Web technologies and geospatial catalogue services.

In the general information domain, much work has been conducted on adding semantics to UDDI and ebRIM. UDDI provides a data structure called TModel that can specify the additional attributes of entities, thus allowing description of the specified ontological concepts. Each service can have one or more TModels that help describe its characteristics. Thus, service capabilities such as function or service input/output can be recorded in the corresponding TModels. Three options for semantic search are available in the UDDI research:

(1) Semantic search functionality is embedded into the registry with some changes to the registry interface to support the semantically augmented query, for example, a RDF representation is embedded in the UDDI query [25];
(2) The functionality is created outside of the registry, without any change to the registry interface [28–30, 45];
(3) The functionality is wrapped as an individual external matching service registered in the registry. In this option, UDDI relays the matching task to the external matching service to enable the different types of matching such as OWL-S, WSDL, and UML [46].

The work described in this paper follows Option 2. It creates semantic middleware in front of the registry, enabling reuse of the service interface from the legacy catalogue. Although the focus is on the CSW-ebRIM profile, the idea of a three-phase service

discovery algorithm from UDDI research [29] was borrowed, and adapted to support the extended information model and process model discovery.

Because the ebRIM information model enables catalogues to handle not only services but also other information resources such as data, it has been adopted by OGC. There have been efforts in the general information domain to add semantics to ebRIM [21–24]. The basic idea is to use those extensibility points such as new kinds of associations, classifications, and additional slots to record corresponding OWL classes, properties and related axioms such as subclassOf. However, few studies of registering OWL-S into ebRIM are available. Although OWL-S is mentioned by Dogac et al. [47], only hierarchical OWL classes addressing service functionalities have been explored for registration in ebRIM. The semantics of service instances such as input and output cannot be used in the search. The system development here focuses on the geospatial domain and explores the registration of semantics for geospatial data, services, and service chains. An important characteristic of the geospatial domain is that an application often includes multiple modeling or processing steps involving large and heterogeneous data volumes. While the Dataset class is a core extension to the OGC-ebRIM profile, the ProcessModel class is a core contribution of this paper in that it addresses the analysis and knowledge sharing demands in the context of the Cyberinfrastructure.

In the geospatial domain, most efforts focus on ontology-based descriptions and semantic matching for discovery of geospatial data and services, with little consideration of the registration of semantics in the catalogue service [9, 48–54]. Our previous work [31, 62] proposes approaches for geospatial service composition using Semantic Web Service technologies and provides a prototype system and platform for service composition. The work in [27] describes briefly the semantics-enabled discovery of geospatial data and services and serves as a basic component in the prototype system implementation. The work here extends previous work by adding process models in the ebRIM information model and provides detailed and comprehensive descriptions of the approach and related work. Several instances of a semantics-enhanced catalogue service have been proposed. Maué [55] proposes a semantic geospatial service catalogue to support the discovery of geospatial services. The service platform proposed by the Open Service Gateway Initiative (OSGI) is used for catalogue implementation. Thus, his implementation does not follow a standard like UDDI or ebRIM. OGC data services, such as those following WFS and WMS specifications, are registered in the catalogue. WSMO is used for a semantic description of geospatial services. The European integrated project ORCHESTRA presents a general abstract specification of a catalogue service [56]. It provides the basic functions for a catalogue service: search and publications. The mappings from itself to implementation specifications such as the OGC catalogue service or UDDI are supported. Although no specific implementation of the semantic extensions of the ORCHESTRA catalogue is provided, this work does provide a concept for a semantic catalogue, which uses ontologies to expand original queries so that semantically related concepts can be used for the catalogue search. Thus, this work provides the guidance for a semantic catalogue at an abstract level. The system described in this paper chooses a catalogue implementation specification as the starting point. It focuses on extensions to the ebRIM information model in the existing CSW-ebRIM standard, with emphasis on the geoprocessing demands in the Cyberinfrastructure.

8 Conclusions and future work

This paper presents an approach to add semantics into current geospatial catalogue services by extending the underlying information model. Semantics for geospatial data, services, and

service chains, represented using OWL/OWL-S, are organized and registered in the CSW by extending ebRIM elements. In particular, the IOPE semantics are introduced for geospatial services and the registration of IOPE semantics is proposed in the CSW-ebRIM profile for loosely-coupled, mixed-coupled, and tightly-coupled services and data. A ProcessModel class is added to the ebRIM model. It is a key extension, addressing the requirements for geospatial information processing and knowledge discovery purpose in a data-rich distributed environment, as opposed to the initially core extension class Dataset in the CSW-ebRIM profile. Middleware to support semantics-enhanced discovery of geospatial data, services/service chains, and process models has been developed. Such middleware can be applied to support materialization of virtual data products.

The approach demonstrates that semantics can be used to improve the data/services/service chains discovery capability of geospatial catalogue services. The process models, working as a kind of geospatial knowledge, can address the analysis issues in the Cyberinfrastructure and support on-demand delivery of geospatial information and knowledge.

The CSW-ebRIM profile has been adopted because it allows catalogues to handle geospatial data, services, and other types of information resources such as applications schemas, software components, and reference documents; also, it can handle process models, as shown in this paper. This capability to register different information resources is demonstrated through the employment of those extensibility points in this paper. There are other catalogue services, for example, the NASA EOS Clearinghouse (ECHO), the U.S. Department of Energy (DOE) Earth System Grid (ESG) Simulation Data Catalogue, discovery frameworks like UDDI, and other application profiles of the OGC catalogue specification like the ISO metadata application profile. Some of them differ in the catalogue query language and communication protocol, while others differ in the information model. If multiple catalogue services must be used, they can be federated to provide a comprehensive discovery of geospatial information. Bai et al. [57] proposes a federation strategy, where a mediator-wrapper based approach is adopted to build a federation service for distinct geospatial catalogues. It is possible to add the semantic middleware in the federation service as a future work, and use the mediated schema approach from conventional data integration research [58] to derive and dispatch the query to other catalogues and assemble the results from multiple catalogue services.

Acknowledgements We are grateful to the four anonymous reviewers, and to Dr. Barry Schlesinger for their detailed comments that helped improve the quality of the paper. This work was funded fully or partially by U.S. NGA NURI program (HM1582-04-1-2021), Project 40801153 supported by NSFC, 863 Program of China (2007AA120501, 2007AA12Z214), LIESMARS and SKLSE (Wuhan University) Special Research Funding.

References

1. Hey T, Trefethen AE (2005) Cyberinfrastructure for e-Science. Science 308(5723):817–821
2. Brodaric B, Fox P, McGuinness DL (2007) Call for special issue on geoscience knowledge representation for cyberinfrastructure, Comput Geosci.
3. Di L (2004) GeoBrain-a web services based geospatial knowledge building system. In: Proceedings of NASA Earth Science Technology Conference 2004. June 22–24, 2004. Palo Alto, CA, USA. 8 pp. (CD-ROM).
4. Papazoglou MP (2003) Service-oriented computing: concepts, characteristics and directions. In: Proceedings of The Fourth International Conference on Web Information Systems Engineering (WISE 2003), Roma, Italy, pp. 3–12.
5. Hendler J (2003) Science and the semantic web. Science 299(5606):520–521

6. Gruber TR (1993) A translation approach to portable ontology specification. Knowledge Acquisition 5 (2):199–220
7. Fonseca FT, Egenhofer MJ, Agouris P, Camara G (2002) Using ontologies for integrated geographic information systems. Trans GIS 6(3):231–257
8. Egenhofer M (2002) Toward the semantic geospatial web. In: The 10th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS), McLean,VA. 4 pp.
9. Lutz M, Klien E (2006) Ontology-based retrieval of geographic information. Int J Geogr Inf Sci 20 (3):233–260
10. W3C (2004) OWL Web Ontology Language Reference, W3C. http://www.w3.org/TR/owl-ref.
11. W3C (2004) Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.
12. Baader F, Nutt W (2003) Basic description logics. In: Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (eds) The description logic handbook. theory, implementation and applications. Cambridge University Press, Cambridge, pp 47–100
13. Martin D, Burstein M, Hobbs J et al (2004) OWL-based Web Service Ontology (OWL-S). http://www.daml.org/services/owl-s/.
14. de Bruijn J, Bussler C, Domingue J et al (2005) Web Service Modeling Ontology (WSMO). http://www.w3.org/Submission/WSMO/ .
15. Akkiraju R, Farell J, Miller JA, Nagarajan M, Sheth A, Verma K (2005) Web Service Semantics—WSDL-S http://www.w3.org/2005/04/FSWS/Submissions/17/ WSDL-S.htm .
16. Farrell J, Lausen H (2006) Semantic Annotations for WSDL (SAWSDL). http://www.w3.org/TR/sawsdl/.
17. Battle S, Bernstein A, Boley H et al (2005) Semantic Web Services Framework (SWSF) Overview, http://www.w3.org/Submission/2005/SUBM-SWSF-20050909/.
18. ISO/TC 211 (2005) ISO19119:2005, Geographic Information—Services.
19. Nebert D, Whiteside A, Vretanos P (eds) (2007) OpenGIS® Catalog Services Specification, Version 2.0.2, OGC 07-006r1, Open GIS Consortium Inc. 218 pp.
20. Martell R (ed) (2008) CSW-ebRIM Registry Service—Part 1: ebRIM profile of CSW, Version 1.0.0, OGC 07-110r2, Open Geospatial Consortium, Inc., 57 pp.
21. Dogac A (ed) (2006) ebXML Registry Profile for Web Ontology Language (OWL), Version 1.5, regrep-owl-profile-v1.5-cd01, Organization for the Advancement of Structured Information Standards (OASIS). 76 pp.
22. Dogac A, Kabak Y, Laleci GB, Mattocks C, Najmi F, Pollock J (2005) Enhancing ebXML registries to make them OWL aware, Distributed and Parallel Databases Journal, Springer-Verlag, July, 18(1):9–36.
23. Liu W, He K, Liu W (2005) Design and realization of ebXML registry classification model based on ontology. In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05), pp. 809–814.
24. Bechini A, Tomasi A, Viotto J (2008) Enabling ontology-based document classification and management in ebXML registries, In: Proceedings of the 2008 ACM symposium on Applied computing, Fortaleza, Ceara, Brazil, pp. 1145–1150.
25. Akkiraju R, Goodwin R, Doshi P, Roeder S (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.
26. OASIS (2004) The UDDI technical white paper, http://uddi.org/pubs/uddi-tech-wp.pdf.
27. Yue P, Di L, Zhao P, Yang W, Yu G, Wei Y (2006) Semantic augmentations for geospatial catalogue service. In: Proceedings of the 2006 IEEE International Geoscience and Remote Sensing Symposium (IGARSS06), 31 July-4 August 2006, Denver, USA. pp. 3486–3489.
28. Paolucci M, Kawamura T, Payne TR, Sycara K (2002) Importing the Semantic Web in UDDI. In: Proceedings of Web Services, E-Business and Semantic Web Workshop. 2002. 12 pp.
29. Sivashanmugam K, Verma K, Sheth AP, Miller JA (2003) Adding semantics to web services standards. In: Proceedings of the 1st International Conference on Web Services (ICWS '03). Las Vegas, Nevada 2003, USA. 7 pp.
30. Srinivasan N, Paolucci M, Sycara K (2004) Adding OWL-S to UDDI, implementation and throughput. In: Proceedings of First International Workshop on Semantic Web Services and Web Process Composition, San Diego, USA 2004. 12 pp.
31. Yue P, Di L, Yang W, Yu G, Zhao P, Gong J (2009) Semantic web services based process planning for earth science applications. International Journal of Geographical Information Science, in press, online accessible. 25 pp. DOI: 10.1080/13658810802032680.
32. Bermudez EL (2004) Ontomet: ontology metadata framework. Ph.D. Dissertation, Drexel University, Philadelphia, USA, 177 pp.

33. OASIS (2005) ebXML Registry information model version 3.0, OASIS Standard, 2 May, 2005. regrep-rim-3.0-os, 78 pp.
34. Martell R (ed) (2008) CSW-ebRIM Registry Service - Part 2: Basic extension package, Version 1.0.0, OGC 07-114r2, Open Geospatial Consortium, Inc., 48 pp.
35. OASIS (2007) Web services business process execution language, version 2.0. Web Services Business Process Execution Language (WSBPEL) Technical Committee(TC).264 pp.
36. Bowers S, Ludäscher B (2004) An ontology-driven framework for data transformation in scientific workflows. In: Rahm E (ed) Proceedings of the international workshop on data integration in the life sciences (DILS 2004), LNCS 2994. Springer, Berlin, pp 1–16
37. Rodriguez MA, Egenhofer MJ (2003) Determining semantic similarity among entity classes from different ontologies. IEEE Trans Knowl Data Eng 15(2):442–456
38. Cardoso J, Sheth A (2003) Semantic e-workflow composition. Journal of Intelligent Info-mation Systems (JIIS) 21(3):191–225
39. Rao J, Su X (2004) A survey of automated web service composition methods. In: Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition (SWSWPC 2004), San Diego, CA, USA, pp. 43–54.
40. Srivastava B, Koehler J (2003) Web service composition—current solutions and open problems. In: Proceedings of ICAPS 2003 Workshop on Planning for Web Services, Trento, Italy, pp. 28–35.
41. Peer J (2005) Web service composition as AI planning—a survey. Technical Report, University of St. Gallen, Switzerland, 63 pp.
42. Wei Y, Di L, Zhao B, Liao G, Chen A, Bai Y, Liu Y (2005) The design and implementation of a grid-enabled catalogue service. In: Proceedings of 25th Anniversary of IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2005), July 25–29, COEX, Seoul, Korea. pp. 4224–4227.
43. ISO/TC 211 (2003) ISO19115:2003, Geographic information—Metadata.
44. Zhao P, Di L, Wei Y (2006) A virtual data product toolkit based on geospatial web service orchestration, Geoinformatics 2006. May 10–12, 2006. Reston, VA, USA.
45. Paolucci M, Kawamura T, Payne TR, Sycara K (2002) Semantic matching of web services capabilities. In: Proceedings of the First International Semantic Web Conference, Sardinia, Italy, June 9–12, 2002, Lecture Notes in Computer Science (LNCS) 2342, Springer, Berlin, Germany, 2002, pp. 333–347.
46. Colgrave J, Akkiraju R, Goodwin R (2004) External matching in UDDI. In: Proceedings of 2004 IEEE International Conference on Web Services, San Diego, USA, 2004. 8 pp.
47. Dogac A, Kabak Y, Laleci GB (2004) Enriching ebXML registries with OWL ontologies for efficient service discovery, In: Proceedings of the 14th International Workshop on Research Issues on Data Engineering: Web Services for E-Commerce and E-Government Applications (RIDE' 04), Boston, USA, pp. 69–76.
48. Klien E, Einspanier U, Lutz M, Hübner S (2004) An architecture for ontology-based discovery and retrieval of geographic information. In: Proceedings of 7th Conference on Geographic Information Science (AGILE 2004), Heraklion, Greece, pp. 179–188.
49. Klien E, Lutz M, Kuhn W (2006) Ontology-based discovery of geographic information services—an application in disaster management. Comput Environ Urban Syst 30(1):102–123
50. Lutz M (2007) Ontology-based descriptions for semantic discovery and composition of geoprocessing services. Geoinformatica 11(1):1–36
51. Kolas D, Hebeler J, Dean M (2005) Geospatial semantic web: architecture of ontologies. In: Proceedings of First International Conference on GeoSpatial Semantics (GeoS 2005). Mexico City, Mexico, Springer, pp. 183–194.
52. Kolas D, Dean M, Hebeler J (2006) Geospatial semantic web: architecture of ontologies. In: Proceedings of 2006 IEEE Aerospace Conference. Big Sky, Montana, March 4–11, 2006. 10 pp.
53. Kammersell W, Dean M (2006) Conceptual search: incorporating geospatial data into semantic queries. In: Proceedings of Terra Cognita 2006, Workshop of 5th International Semantic Web Conference. November 5–9, 2006. Athens, Georgia, USA. 10 pp.
54. Lutz M, Kolas D (2007) Rule-based discovery in spatial data infrastructures. Transactions in GIS, special issue on the geospatial semantic web 11(3):317–336
55. Maué P (2008) An extensible semantic catalogue for geospatial web services. International Journal of Spatial Data Infrastructures Research 3(2008):168–191
56. Hilbring D, Usländer T (2006) Catalogue services enabling syntactical and semantic interoperability in environmental risk management architectures. Proceedings of the 20th International Conference on Informatics for Environmental Protection (EnviroInfo 2006), September 6–8, 2006, Graz, Austria. pp. 39–46.
57. Bai Y, Di L, Chen A, Liu Y, Wei Y (2007) Towards a geospatial catalogue federation service. Photogramm Eng Remote sensing 73(6):699–708
58. Halevy AY (2001) Answering queries using views: a survey. VLDB J 10(4):270–294

59. Russel S, Norvig P (2003) Artificial intelligence: a modern approach, 2nd edn. Prentice-Hall Inc, USA, pp 375–458
60. Cyberinfrastructure Council (2007) Cyberinfrastructure Vision for 21st Century Discovery, National Science Foundation, USA, March, 2007, 64 pp.
61. Foster I, Kesselman C, Tuecke S (2001) The anatomy of the grid: enabling scalable virtual organizations. International Journal Supercomputer Applications 15(3):200–222
62. Yue P, Di L, Yang W, Yu G, Zhao P (2007) Semantics-based automatic composition of geospatial Web services chains. Comput Geosci 33(5):649–665
63. Fuger S, Najmi F, Stojanovic N (eds) (2005) ebXML Registry Information Model Version 3.0, regrep-rim-3.0-os, Organization for the Advancement of Structured Information Standards (OASIS). 78 pp.

Dr. Peng Yue holds a Ph.D. in GIS from the Wuhan University (2007). He is currently an associate professor at State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing of Wuhan University, China. His research interest is on GIS interoperability, Web GIS, and Geospatial Semantic Web. He has been involved in many related research projects, including Choreographed Intelligent Web Services for Automated Geospatial Knowledge Discovery funded by U.S. NGA NURI, GeoBrain project funded by U.S. NASA REASoN program, Metadata Tracking in Geospatial Service Chaining and Geospatial Data Provenance funded by NSF of China, Grid GIS and Semantic Web-based Intelligent Geospatial Web Service funded by Ministry of Science and Technology of China.

Dr. Jianya Gong is the professor and director of the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China. He studied as a PhD candidate at Wuhan Technical University of Surveying and Mapping and the Technical University of Denmark from 1988–1992 and received his Ph.D. in 1992. His research interests include geospatial data structure and data model, geospatial data integration and management, geographical information system software, geospatial data sharing and interoperability, Photogrammetry, GIS and remote sensing applications.

Dr. Liping Di holds a Ph.D. in Remote Sensing and GIS from the University of Nebraska-Lincoln. He is currently the professor and director of the Center for Spatial Information Science and Systems (CSISS) (formerly LAITS), George Mason University. His research interest is in the area of GIS, remote sensing, interoperability, Semantic Web, global climate and environmental changes.

Lianlian He holds a B.Sc. degree in Informatics and Computational Sciences (2002) and a M.Sc. degree in Computational Mathematics from the Wuhan University (2005). She is currently working as a lecturer in the department of Mathematics, Hubei University of Education, China. Her research interest is on the applications of computational mathematics methods in Geoinformatics and Bioinformatics.

Dr. Yaxing Wei holds a Ph.D. in Safety Engineering from the University of Science and Technology of China. He was a visiting research assistant at the George Mason University and is currently the post-doctoral research associate at Environment Science Division, Oak Ridge National Lab, United States. His research interest is in the area of Web GIS, geospatial data management and dissemination, global climate change, and ecosystem modeling.