Computers & Geosciences 37 (2011) 1752–1762

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Semantic-based web service discovery and chaining for building an Arctic spatial data infrastructure

W. Li [a,*], C. Yang [a,*], D. Nebert [b], R. Raskin [c], P. Houser [a], H. Wu [a], Z. Li [a]

[a] Department of Geography and GeoInformation Science and Center of Intelligent Spatial Computing for Water/Energy Science, George Mason University, Fairfax, VA 22030, USA
[b] Federal Geographic Data Committee, 590 National Center, Reston, VA 20192, USA
[c] NASA Jet Propulsion Laboratory, Pasadena, CA, USA

ARTICLE INFO

Article history:
Received 17 May 2010
Received in revised form
3 June 2011
Accepted 8 June 2011
Available online 23 July 2011

Keywords:
SDI
Hydrology
Arctic
Ontology
Crawler
Semantic
Service chain
Knowledge base

ABSTRACT

Increasing interests in a global environment and climate change have led to studies focused on the changes in the multinational Arctic region. To facilitate Arctic research, a spatial data infrastructure (SDI), where Arctic data, information, and services are shared and integrated in a seamless manner, particularly in light of today's climate change scenarios, is urgently needed. In this paper, we utilize the knowledge-based approach and the spatial web portal technology to prototype an Arctic SDI (ASDI) by proposing (1) a hybrid approach for efficient service discovery from distributed web catalogs and the dynamic Internet; (2) a domain knowledge base to model the latent semantic relationships among scientific data and services; and (3) an intelligent logic reasoning mechanism for (semi-)automatic service selection and chaining. A study of the influence of solid water dynamics to the bio-habitat of the Arctic region is used as an example to demonstrate the prototype.

1. Introduction

The Arctic region has had the greatest climate change impact observed over the past century (ACIA, 2004; White et al., 2007). For example, in the past 20 years, the rate of the melting of sea ice in the Arctic Ocean has increased rapidly (Wadhams and Munk, 2004; Barber et al., 2008). Observations also show that the extent and depth of snow cover on land and ice as well as the size of the ice sheets on small glaciers in the Arctic are decreasing (Notz, 2009). If this trend continues, the Arctic Ocean will be completely ice free during the summer as early as 2050 (Flato and Boer, 2001; Barber and Massom, 2007). The loss of solid water resources leads directly to the rising of sea level and the endangerment of the habitats of polar life. Providing accurate monitoring of the large-scale dimensions of solid water concentrations in the high latitudes of the Northern Hemisphere becomes a crucial task for Arctic scientists. Empirical and modeling studies have demonstrated the influential role of solid water resources to biological, chemical, and geologic processes within the global heat budget (Madsen et al., 2007; Roesch et al., 2001).

In the 1980s, scientific data relating to the above studies were captured, summarized, and shared through paper media (Hey and Trefethen, 2005). In the 1990s, the emergence of the World Wide Web and Cyberinfrastructure (CI), which integrates hardware, digitally enabled sensors, and an interoperable suite of software and middle services and tools, transformed how scientists, educators, government officials, and the public exchanged ideas and shared knowledge (Holzmann and Pehrson, 1994; NSF, 2003; Yang et al., this issue). However, huge volumes of rapidly expanding data and ever-changing experimental and simulation results are largely disconnected from each other due to the distributed nature of the communities which create them. In addition, heterogeneity is a ubiquitous problem when exchanging and sharing these resources. Although the advancement of open standards and available transformation tools greatly improve the syntax-level interoperability (Goodchild et al., 1999; Johnson et al., this issue), the structural and semantic heterogeneity still presents significant impediments. Moreover, traditional simulation models normally only include local datasets rather than distributed data and processing services (Díaz et al., 2008); hence, the extensibility and the capability to provide an integral study are limited.

Another challenge in utilizing the resources for scientific modeling is that precise predictions are difficult due to fundamental and irreducible uncertainties (Dessai et al., 2009). In Arctic climate studies, e.g., analyzing and predicting habitat alterations caused by the melting of snow or sea ice, uncertainties could

* Corresponding authors. Tel.: +1 7039934742; fax: +1 7039939299.
E-mail addresses: wli6@gmu.edu (W. Li), cyang3@gmu.edu (C. Yang).

0098-3004/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2011.06.024

originate from limitations in scientific knowledge (e.g., ice sheet dynamics) or human activities (e.g., greenhouse gas emissions). These uncertainties create a theoretical limit of predictability (LOP), as shown in the red curve in Fig. 1, meaning that the prediction error (Y axis) increases with increasing forecasting time (X axis). Practically, some randomness (e.g., the chaos in the Earth and environmental system) leads to fluctuations around the LOP, which forms a wave-like Real-LOP. Even in the most appropriate model, e.g., Model A in Fig. 1, which minimizes error caused by structural complexity (Melbourne and Hastings, 2009), a more precise analysis and prediction can be generated by inputting data with less noise (Data A in Fig. 1). Usually, scientists are limited to the use of datasets that are available to them. They often have little knowledge of the existence and location of datasets that could be a better fit for their model (Gray et al., 2005; Singh, 2010; Tisthammer, 2010).

Scientists face technical challenges when implementing an interoperable geospatial cyberinfrastructure (GCI; Yang et al., 2010) to facilitate the discovery, federation, and seamless fusion of scientific data from disparate and distributed resources. A spatial data infrastructure (SDI), aiming to address this challenge, is to facilitate and coordinate the exchange and sharing of geospatial data between stakeholders in the spatial data community (Nebert, 2004; Rajabifard et al., 2002). SDI research encompasses the policies, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure the availability and accessibility to the spatial data (Holland et al., 1999) (As this paper focuses on providing a technical solution for building an ASDI, policy-related issues will not be emphasized.) Over the past decade, many SDIs have been built at a national or regional scale (Bernard et al., 2004; Coleman and Nebert, 1998; Craglia and Annoni, 2006; Georgiadou et al., 2005; Jacoby et al., 2002; Parent et al., in this issue; Rajabifard et al., 1999). However, very few SDIs were specifically developed to support Arctic research. This paper reports our research and development using spatial web portal (Yang et al., 2007) toward building an ASDI where Arctic geospatial data, information, and services are shared and chained in a seamless manner for effective Arctic hydrology studies and decision making. Several key issues must be addressed on how to (1) enable the automatic discovery of spatial data that exist in a distributed computing environment; (2) build a knowledge base to improve machine understanding of the data and services discovered; and (3) provide an intelligent resource search mechanism to implement on-the-fly service chaining for decision making. Here to connect various services in a "chain" means to composite services that implement different functions into an integral service in a certain order (Nadine, 2003).

Section 2 introduces the key components of an ASDI; Section 3 proposes the methodologies to enhance the capabilities of ASDI, including a hybrid approach to support Arctic Web service discovery, a semantic-enabled search service for query interpretation, and a service chaining model to support decision making; Section 4 demonstrates a proof-of-concept ASDI prototype that implements the aforementioned approaches; and Section 5 concludes with remaining issues and future research directions.

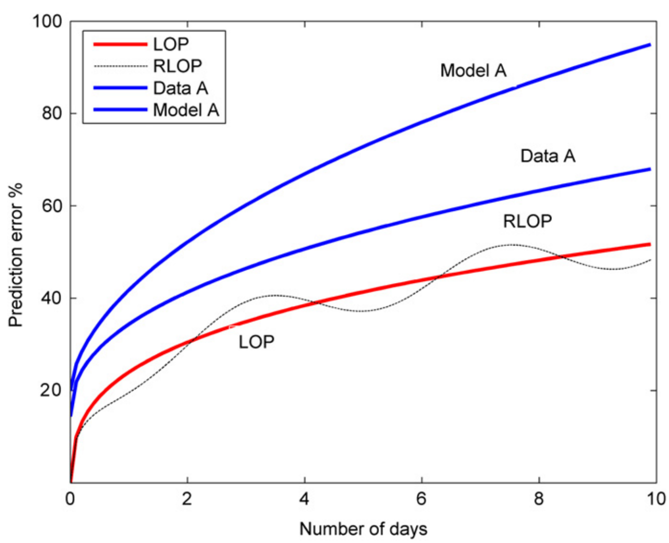

Fig. 1. Prediction error as a function of forecast time (adapted from Suranjana and Van Den Dool, 1988). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Key components of an ASDI and a use case

The ASDI research can be traced back to 2001, when the Arctic Research Consortium of the U.S. in the white paper "ASDI for

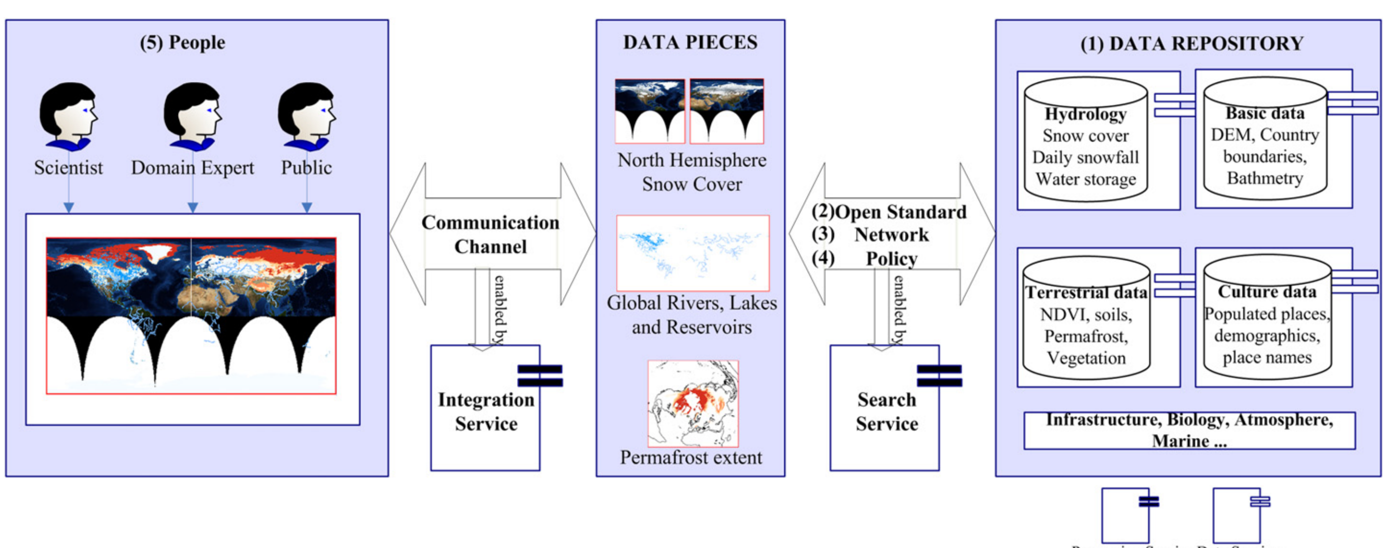

Fig. 2. Key components of ASDI.

Scientific Research" conceived the important components in developing an ASDI (Sorensen et al., 2004). In 2007, the First International Circumpolar Conference on Geospatial Sciences and Applications in Canada, coordinated and encouraged the eight Arctic circumpolar countries to move toward a common ASDI (Huebert, 2009). This ASDI is built based on the SDI concept, and would provide more efficient integration and a more robust management of Arctic data, because Arctic data are often lost within the broader context due to their "polar" nature. The proposed ASDI, adopted from a general SDI (Rajabifard et al., 2002), includes the same five key components (Fig. 2): (1) a data repository stores and manages the Arctic science data, including the attribute data and metadata. (2) Open standards make the scientific data from diverse data sources interoperable. The original data are reproduced and delivered on demand through standardized web services, such as OGC Web Map Service (WMS) (de La Beaujardiere, 2004), OGC Web Feature Service (WFS) (Vretanos, 2005), or OGC Web Coverage Service (WCS), (Whiteside and Evans, 2006) or some specialized distribution services (Zhou et al., this issue). (3) Clearinghouse network functions as the main communication medium which enables interaction between data producers and data consumers. (4) Policies are used to regulate data access and licensing, protect the privacy of Arctic data, and provide custodianship at all administrative levels. (5) People, such as Arctic scientists and domain experts, define their resource needs. They are the stakeholders of transaction processing and decision-making. In addition to the static components, the ASDI also provides an automatic discovery mechanism to provide access to all available data, and a search and integration service to analyze and visualize the data.

Using a spatial web portal (Yang et al., 2007), an ASDI is designed to enable an integrated science analysis environment. For example, to study the influence of melting snow and sea ice on habitat changes of polar wildlife, a scientist "enters" the ASDI and the following scenario unfolds:

1. Initially, an automated service in the SDI discovery identifies all the distributed Arctic data resources (latitude [70, 90]; longitude [−180, 180]) across a public network and places them in a virtual data repository. We called the repository "virtual" because it does not actually include real data, instead only metadata and online address of the real datasets are stored. The scientists, who are the data consumers, are only presented with a transparent search interface which hides all implementation details.
2. A scientist opens the user interface of the ASDI and assembles the query elements required to narrow down the available information. The query elements should be geophysical parameters, such as snow cover, sea ice concentration, precipitation, and biodiversity.
3. The search request is handled by the "Search Service" and passed to the virtual repository. Through "smart" processing (discussed in Section 3.3), the most relevant datasets are discovered.
4. Results from disparate sources are collected and returned to the scientist. The response contains a list of relevant datasets described by a full representation of the metadata, including the data format, temporal and spatial coverage data, and data projections. Then an "Integration Service" mosaics the data if a single dataset cannot cover the spatial extent for the analysis, and integrates all of the needed datasets and visualizes the result. This process includes some automation for the complex operations so that the scientist can focus on analyzing observations and findings rather than the technical details.
5. Examining the results composited, the scientist decides whether to acquire more datasets for further analysis. By clicking the embedded URLs, the scientist gains direct access to the needed resources to feed a simulation model or conduct a cross-correlation of multiple variables.

3. Methodology

In this section, we discuss the logic that enhances the capabilities of the ASDI to provide an automatic discovery mechanism to collect distributed data (Section 3.1), the buildup of a hydrology ontology to model the latent semantic relationship among the data (Section 3.2), and a smart search and integration service to chain and visualize the datasets to enable a semiautomatic science workflow (Section 3.3).

3.1. Data discovery: A hybrid approach for Arctic science using virtual repositories

The scientific data and research findings for Arctic studies are distributed worldwide on the open and dynamic Internet. Making all relevant data from all possible sources available to researchers has the potential to revolutionize how science is conducted (Fox et al., 2006). Among several possible approaches for scientific data and service discovery, e.g., Internet server, general search engines, Z39.50 (Lynch, 1991), Universal Description Discovery and Integration (UDDI, Clement et al., 2005), Open Archives Protocol for Metadata Harvesting (OAI-PMH, de Sompel et al., 2004), the most common is a centralized catalog with registered metadata of distributed datasets (Ma et al., 2006; Nogueras-Iso et al., 2005; Singh et al., 2003). Especially after the adoption of the OGC Web Catalog Service (CSW; Nebert and Whiteside, 2005), OGC CSW-based web catalogs have become the major mechanism to support multiple users in discovering relevant scientific data and services from heterogeneous and distributed repositories.

The CSW-based web catalog helps users to discover data; however, each CSW focuses on one specific topic, which is insufficient for a comprehensive multidisciplinary study, needed in an Arctic study. The catalog approach is based on the premise that data producers have registered their data into the catalog with the correct classifications. This assumption is sometimes not met because some providers do not register their data into the catalogs or publish their data through other approaches (Al-Masri and Mahmoud, 2007). To utilize the advantages of the CSW-based discovery mode and to gather all available resources for Arctic research, we built a geo-bridge that connects with distributed CSW catalogs to discover existing registered resources, and a web crawler (Li et al., 2010) to actively search for data dispersed on the Internet but not registered in the CSW catalogs.

Fig. 3 demonstrates the workflow service enabled by the hybrid approach—using both multicatalogue searching (left side of Fig. 3) and active crawling (right side of Fig. 3) for automatic data collection. The geo-bridge distributes the data discovery task to the multicatalogues and the entry of the active crawler. For catalog discovery, a query container will collect the search criteria from a template-based GUI and translate the query into a Common Query Language (CQL) or OGC Filter specification-encoded XML, which is compliant with CSW implementation specifications. By feeding the query into multiple known catalogs using Asynchronous JavaScript and XML (AJAX) technique, the total time cost is reduced from the sum of the catalog connection times to the maximal response time among all the connections, and the discovery process is significantly accelerated (Li et al., in press). As different catalogs may choose different metadata profiles, extending the capability from parsing one specific standard, e.g., Dublin Core (DC; Weibel et al., 1998), to adapting it to understand all popular profiles is of great importance. In the ASDI, an XML response parser selects one of the interpreters of metadata based on standards, such as the FGDC CSDGM (FGDC, 1998), DC, ISO 19139 (ISO, 2007), ISO 19119 (ISO, 2001), or OASIS ebRIM (OASIS, 2002) to resolve metadata on-the-fly from a tree-based response file.

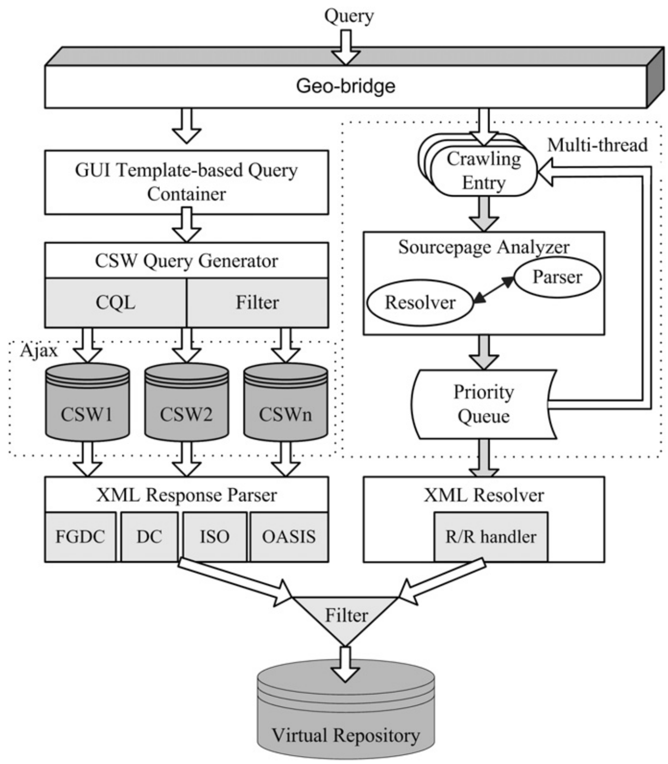

Fig. 3. Data collection for building the virtual repository.

In addition to the multicatalogue search, an active crawler was developed to discover dispersed services that are not registered in the web catalog (Li et al., 2010). The discovery of WMS is currently supported by the crawler because: (1) WMS is the most popular web service standard in enabling geospatial interoperability. The number of existing WMSs is much more than other OGC services, such as WFS and WCS which deliver vector data and allow more complicated scientific inquires. As WMS, WCS, and WFS all follow the same requesting procedure, it is easy to adapt the crawler to discover other existing data and services once the feasibility of the approach is verified by using WMS as a case study. (2) Currently, many scientific data providers, e.g., NSIDC, NASA, and Norway's mapping agency all publish their scientific data into WMS because it provides an easy way to visualize and integrate scientific data. This procedure acts as a quick preview of the potential scientific analysis that can be conducted. Scientists can eventually access the raw data and feed them into various models from the metadata of the discovered WMSs. The active crawler starts with one seed URL and then crawls the web to find a hyperlink, indicating a WMS and subsequently parses it with a WMS Capabilities analyzer. The buffer selectively caches web source codes linked by URLs. A source-page analyzer is used to analyze the web source code which has been cached in the buffer. It extracts all out-links and transforms the relative URLs of the links into absolute URLs. After completing the process, the analyzer removes the source code based on the strategy adopted for the buffer module. The priority queue is used to store the crawled webpages in descending order of their possibility to contain or link to data. Once a dataset is found, all metadata information will be extracted and handled by an XML resolver. The crawling process is improved by a multithreading strategy, which provides a speed gain of up to 10 times relative to using a single thread (Li et al., 2010).

New datasets are continually discovered from catalogs and by the active crawler. All the metadata of the datasets are put into the virtual repository with repeated information filtered out.

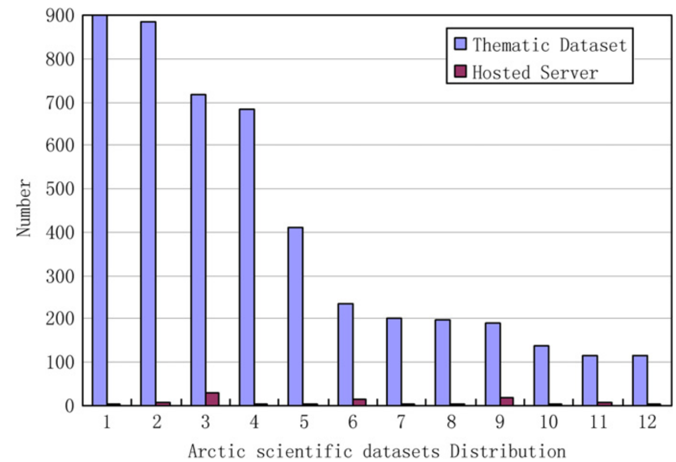

Fig. 4. Arctic data distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

There are 12,123 thematic datasets covering the Arctic area that have been found from distributed online resources. Fig. 4 shows the top 12 Arctic data providers, which are: (1) Government of Canada (7147), (2) Greenland Digital Atlas (884), (3) NOAA (716), (4) Canada Geobase (684), (5) Alaska Mapped and the Statewide Digital Mapping (409), (6) Geological Survey of Norway (235), (7) Norkart Geoservices (203), (8) National Snow and Ice Data Center (198), (9) Norwegian Water Resources and Energy Directorate (190), (10) NASA (137), (11) The Norwegian Forest and Landscape Institute (117), and (12) Norwegian Meteorological Institute (116). From Fig. 4, we can also tell the number of web servers on which the thematic datasets reside (red bar). Although the services are widely distributed around the world, on average, there are more than 100 ($\sim$114) data layers clustered on a single web server. This distribution pattern verified the clumped distribution of online spatial web services (Li et al., 2010).

3.2. Buildup of a domain knowledge base to model Arctic hydrological knowledge

Once rich resources have been collected and stored in the ASDI repository, there is a need to provide a mechanism to help scientists find the most suitable data to perform automated analyses for their study. One problem is the semantic heterogeneity of the data and services. The variability of service sources may cause different terms to refer to the same information in different datasets (synonym problem) (Nauman et al., 2008). It is also possible for different services to use the same term when referring to different types of information (polysemy problem) (Nauman et al., 2008). For example, the term "Creek" may refer to a small stream or an inlet/narrow cove of the sea. Users in different contexts or with different needs, knowledge, or linguistic habits will describe the same information using different terms. Indeed, previous research found that the degree of variability in descriptive term usage is much greater than commonly suspected (Deerwester et al., 1990). One study shows that the probability of two people choosing the same keywords for a single well-known object is less than 20% (Furnas et al., 1983). Such semantic heterogeneity problems can cause serious conflicts during the selection of suitable data sources within service chaining processes (Bernard et al., 2003). To solve this problem and advance the Arctic climate change and hydrological research, we: (1) developed a domain knowledge base (ontology) to disambiguate between different terminologies by explicitly defining a conceptualization and (2) provided an intelligent search tool for resource selection in semantic-enabled service chaining.

A knowledge base, also called an ontology, encodes the explicit logic definition for a shared conceptualization (Gruber, 1993). The use of an ontology helps to discover the implicit relations between concepts that are not usually made explicit in traditional databases (Latre et al., 2009). Several hydrology ontologies have been developed to serve the needs of the broader hydrology community as well as Arctic researchers. The Consortium of Universities for Advancement of Hydrologic Science (CUAHSI) developed a taxonomy-based ontology. It focuses on classifying the key components (e.g., precipitation, radiation, and water body) in the water cycle and the interaction of the hydrosphere with the atmosphere and the biosphere (Tarboton et al., 2006; Maidment, 2008). Another source of hydrological knowledge is derived from NASA's Global Change Master Directory (GCMD) (Olsen, 2001) keyword collection. The GCMD contains 1000 controlled keywords used by clearinghouses to classify resources in the Earth sciences. An additional 20,000 uncontrolled keywords in climatology, marine, geology, etc. were extracted from the descriptions of the data and service providers. Other existing data dictionaries for environmental knowledge modeling include the General Multilingual Environmental Thesaurus (GEMET, 1999) and INSPIRE hydrological theme initiatives (Latre et al., 2009).

The taxonomy and controlled keywords provide valuable guidance to differentiate terms; however, these efforts contain few interrelation and association definitions. To overcome this issue and make the available terminologies maximally reusable, we incorporate the infrastructure of the largest Earth science ontology, Semantic Web for Earth and Environmental Terminology (SWEET) (Raskin and Pan, 2005), to build the hydrology ontology for Arctic studies. Fig. 5 shows the modularized design of the ontology. SWEET 2.0 builds on basic math, science, and geographic concepts to include additional modules for the planetary realms, such as Hydrosphere, Cryosphere, Atmosphere, Geosphere, Biosphere, etc. This modularized design facilitates the domain specialists to build self-contained specialized ontologies that extend existing ones.

Based on the SWEET 2.0 infrastructure, we built our hydrology ontology to express both human- and machine-understandable facts and logics. The upmost layer of Fig. 6 (Li, 2010) demonstrates an ontology fragment containing the knowledge that will be used for logic inference. Building such an ontology requires the following steps. (1) Conceptualization: we extracted and combined all of the hydrology-related terminologies from CUAHSI-, GCMD-, GEMET-, and INSPIRE-controlled vocabularies to form a comprehensive vocabulary. (2) Facet mapping: as a hydrology ontology module of SWEET 2.0, the inside design of the module classifies the terminologies into several facets (Space, Property, Substance, Planetary Realm, and Phenomena) as shown by the colored nodes in Fig. 6. (3) Abstraction of class: abstraction is an important approach to model the world, as shown in Fig. 6 where each node represents a class. (4) Building the interrelationship: the ontology goes beyond a simple "is–a" classification tree to model the connections of classes across facets. These connections or interrelations are represented by the arrow links shown in Fig. 6. (5) Domain ontology models: once the conceptual model is created, we encode it in a machine-readable format. For this purpose, we use the W3C standardized languages Resource Description Framework (RDF) (Brickley and Guvha, 2004) and Web Ontology Language (OWL) (Dean and Schreiber, 2004). Both languages are based on formal semantics and are serialized and exchanged using XML. RDF and OWL describe the model in triple statements ⟨Subject, Predicate, Object⟩. For example, "Ice" can be measured by a parameter "Ice Concentration," so the statement is mapped as ⟨"Ice," "measuredBy," "Ice Concentration"⟩. Compared with RDF, OWL has a stronger syntax and vocabulary to restrict the flexibility in logic representation. However, these restrictions make the machine reasoning more decidable than using RDF alone. Therefore, we use OWL to model our domain ontology.

3.3. Semantic reasoning and chaining for an integral science study

Semantic reasoning is the core component of the semantic search engine. Given a user query, syntax analysis, semantic analysis, and information retrieval tasks from a heterogeneous environment are performed in sequence (Li et al., 2008). Syntax analysis focuses on analyzing components of a query sentence, as

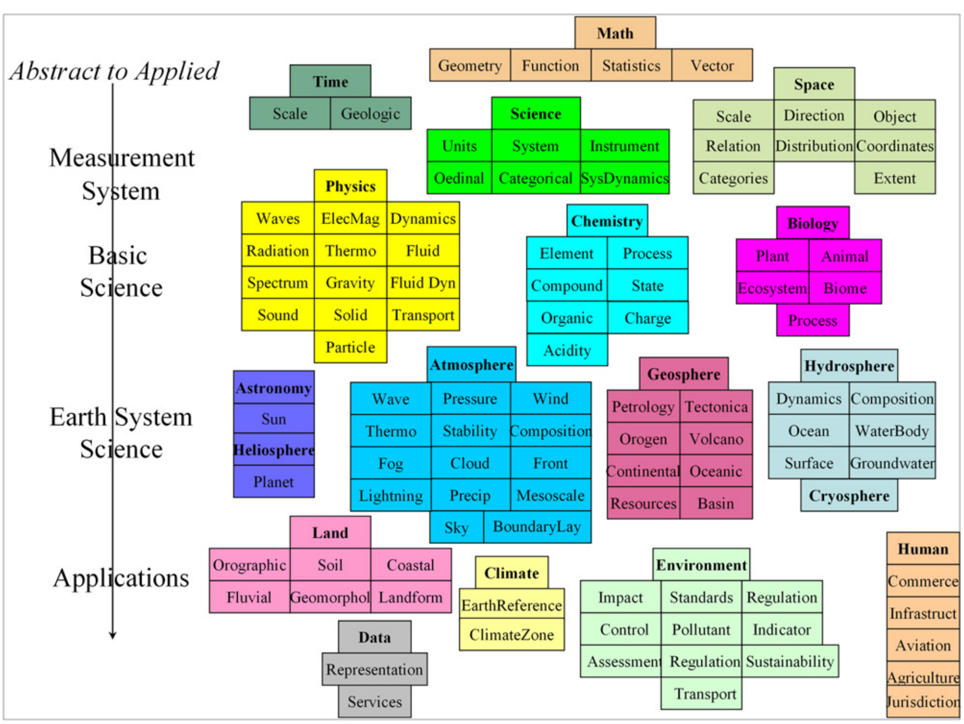

Fig. 5. SWEET 2.0 ontology.

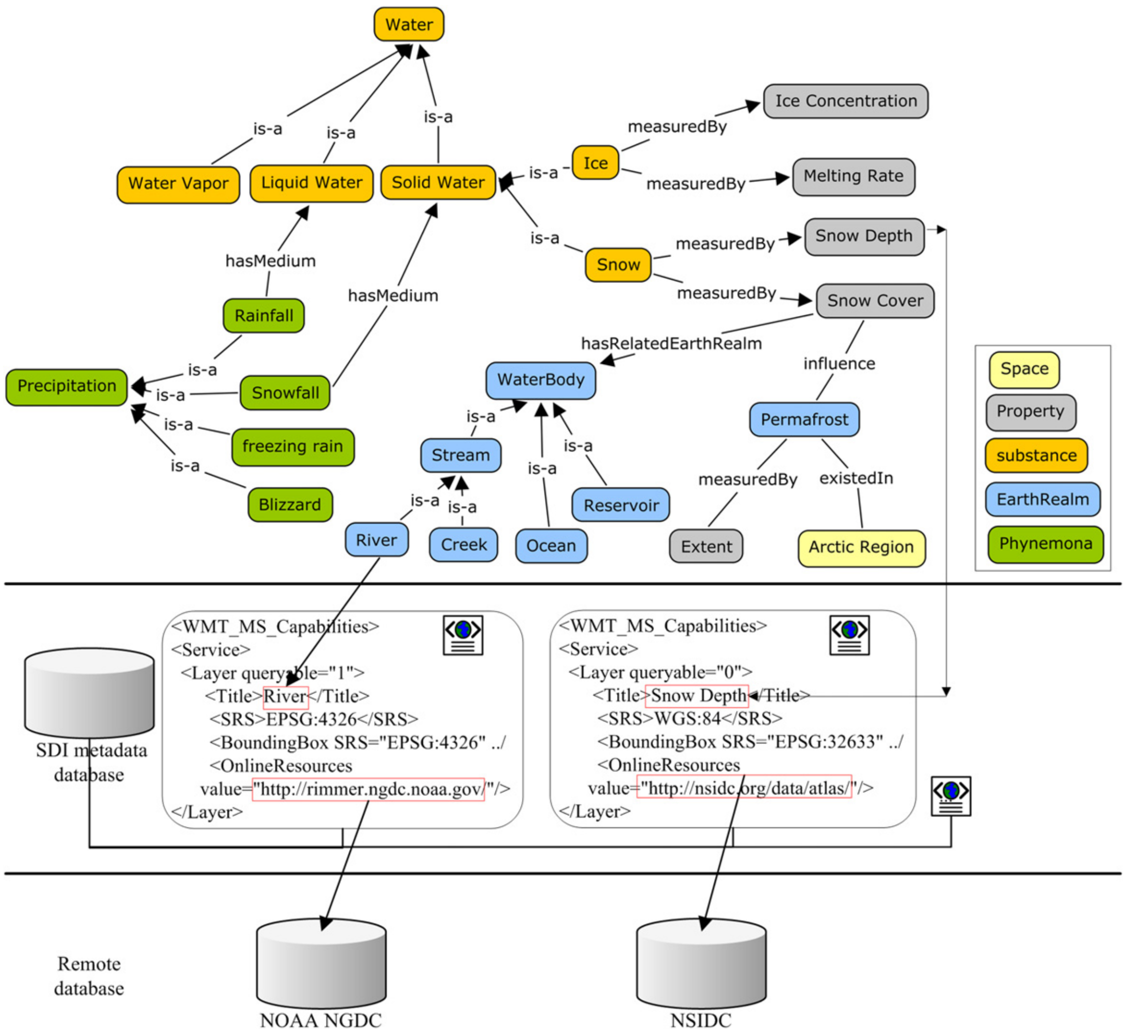

Fig. 6. Ontology fragment and its linkage to metadata and the real science data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

well as getting the "central word," i.e., the exact object in which the user is interested. This would help us to efficiently retrieve the ontology and proper candidates. Syntax analysis can be conducted by either providing a query template for a user to map phrases into different dimensions, e.g., "WHAT," "HOW," "WHEN," and "WHERE," provided in a GUI, or relying on a natural language parser, such as the Stanford open-source statistical parser (Klein and Manning, 2003). In this paper, a template-based approach is chosen to save client parsing time. After syntax analysis, a user query is mapped into two query levels: logic and formal, for semantic reasoning. When reasoning is conducted, complex queries will be decomposed into subqueries. For example, suppose a researcher wants to study: "how does solid water melting influence stream flow in the Arctic Region over the summer?" Through syntax analysis, "solid water" can be distinguished as an event, which is the central word of the whole sentence, "melting" as the state change process, "Arctic" as a place, and "summer" as a time. Given this information, the natural language query can be transformed to a Description Logic (DL)-based query for machine reasoning:

$$\begin{aligned} Q1: \; & \text{Solid Water} \cap \exists\, \textit{hasProperty}.\text{Melt} \cap \exists\, \textit{hasObject}.\text{Stream} \\ & \cap \forall\, \textit{takePlaceIn}.\text{Arctic} \forall \textit{hasTime}.\text{Summer} \end{aligned}$$

In which, $\cap$ is used to represent conjunction. A statement, e.g., $A \cap B$, is true only when both A and B are true. If either A or B is false, the statement will be false. $\exists$ (can be read as "there exists") is the existence quantifier that expresses partial relationships. The universal quantifier $\forall$ (can be read as "for all") is used to formalize that a statement is true for everything.

By iteratively unfolding Q1 from the central word "Solid Water" expressed by the above notation, more useful information can be retrieved based on the knowledge encoded in the ontology. This process is called query decomposition. Given the ontology fragment provided in Fig. 6, Q1 could be decomposed into:

$$Q1a: \; \text{SomeSWClass} \cap \exists \textit{isSubClassesOf}.\text{Solid Water}$$

$$\begin{aligned} Q1b: \; & (\text{AProperty} \cap \exists\, \textit{isSubClassOf}.\text{Property}) \\ & \cap (\text{AProperty} \cap \exists\, \textit{isPredicateOf}.\text{SomeSWClass}) \\ & \cap\; (\text{Parameter} \cap \exists\, \textit{isObjectOf}.\text{SomeSWClass}) \end{aligned}$$

$$Q1c: \; \text{SomeStreamClass} \cap \exists\, \textit{isSubClassesOf}.\text{Stream}$$

$$\begin{aligned} Q1d: \; & (\textit{Parameter} \cup \textit{SomeStreamClass}).\text{hasData} \\ & \cap \forall\, \text{takePlaceIn}.\textit{Arctic} \;\forall\, \text{hasTime}.\textit{Summer} \end{aligned}$$

In this query, Q1a and Q1c aim to find $\langle n_1, n_2 \ldots n_k \rangle$ of all of the subclasses and other related terminologies of the given terms. This type of inference could be considered as a query expansion process on the class level, because terminologies which have similar meanings but are not designated as a query term could be provided.

From the ontology provided in Fig. 6, {"snow" and "ice"} will be returned as the set of "SomeSWClass" for Q1a and {"River" and "Creek"} will be returned as the set of "SomeStreamClass" for Q1c. Q1b is formed by checking all of the roles that are connected with "SomeSWClass" and the connected predicate is a type of "Property." "AProperty" is the intermediate variable and "Parameter" is the variable for the expected results. Given the knowledge base above, "Ice Concentration," "Snow Cover," and other parameters that are used to measure the variation of snow and ice are returned. Compared with Q1a, Q1b, and Q1c, which infer results within the scope of the knowledge base, Q1d can be considered as an external search. After the desired variables in Q1a–c have been inferred, Q1d redirects the query request with the values of desired variables as the keywords to the repository of ASDI. Through matching the service metadata (middle layer in Fig. 6) with given keywords, all relevant services that encapsulate the data are discovered. Following the metadata records, the URL that indicates the online resources of the actual scientific data is also found (lowermost layer in Fig. 6). This way, not only the data sources that are semantically related to their interests are provided but also the bridge between scientific data and scientific processes are well established.

The inference engine adopted was the Jena Semantic Web Framework for Java (Carroll et al., 2004). The processing procedures are as follows. (1) Ontology is loaded into the memory or persistent storage maintained by Jena. (2) The DL-based queries are transformed into formal SPARQL (Prud'hommeaux and Seaborne, 2005) queries. (3) Through the Jena query API, the subqueries are conducted and results are retrieved. (4) Query results are combined to obtain expanded and more specific information. The implicit association inference is enabled by recursively traversing the ontology tree to which the query term belongs. This relation is important because the associations of a class should contain both its own associations and its ascendants' associations.

With the wide availability of distributed web services for scientific data and discovery mechanisms described above, the services can be chained to create a complex scientific workflow. To connect and compose these web services, input and output data types must be aligned and data must be transformed as the services are executed (Bowers et al., 2004). Service chaining can be categorized into a centralized control flow pattern and a cascaded control flow pattern (Alameh, 2002). A centralized control flow pattern requires the chaining engine to have prior knowledge of each service and the types of intermediate results produced in a serial chain. This pattern only implements the lowest level of automation. The cascaded control flow pattern uses a backward chaining approach to break the goal into subgoals recursively until all levels of subgoals are satisfied by certain services. If the services requested are not available, the ASDI provides seamless connection with other popular geo-catalogs, e.g., GEOSS and GOS catalog, to discover the needed services from remote repositories. If there are still missing services at this point, the chaining procedure will convert to the integration procedure by compositing other available services in the chain. In comparison to centralized service chaining, cascaded chaining improves the level of flexibility but also introduces design complexity. In our implementation, we have chosen cascaded chaining for the data discovery and retrieval process, whereas the services for visualizing the results are controlled by the client. This approach makes the discovery process reusable due to the fact that different applications which use other ways to visualize the data can use the same service chain that was previously proposed. In addition, this particular service chain can act as a part of other chains or applications (Friis-Christensen et al., 2009).

As Fig. 7 shows, the mediation service is composed of two services: search and integration. The search service is responsible for locating the most needed data or services and the integration service is in charge of the seamless integration and composition of identified services. This service can be further broken down into: (1) query decomposition service; (2) reasoning service; (3) content-matching service; (4) portrayal service; (5) reprojection service; and (6) overlay service. The strict order of sequences are that (1) should occur before (2), because (2) uses (1)'s output as input; (2) occurs before (3), because after the querying the KB and all the relevant keywords are obtained, the actual metadata

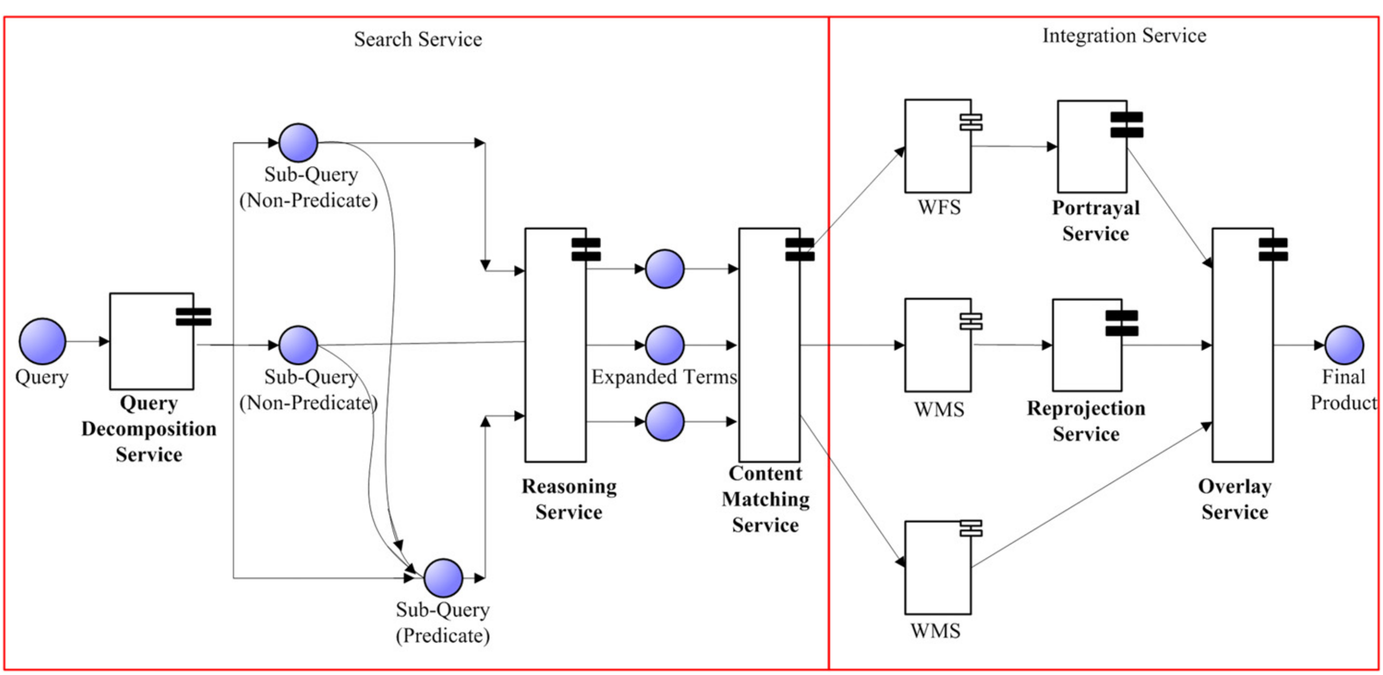

Fig. 7. Service chaining scenario for an integral study (blue nodes represent input, intermediate results, and output). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

matching is conducted; (1)–(3) come before (4) and (5), because (4) and (5) are to conduct format/projection transformation after the matched datasets are returned from (3); and (4) and (5) come before (6). (6) should be conducted when all the data preparation is finished; therefore, it should be the last service in the chain.

As the entry point of the service chain, the query decomposition service and the reasoning service are enabled by semantic technologies, without which the query would be directly sent to a content-matching service. A query decomposition service will conduct nonpredicate queries (such as Q1a and Q1c) first and then predicate (such as Q1b) queries. This sequence is based on the following assumption: the associations (represented by predicate) that the parent class has could be inherited by the child classes. Therefore, by retrieving more event-like classes by a nonpredicate query and feeding them into the predicate query, more candidate results will be retrieved. The expanded subqueries are sent to the reasoning service and all relevant keywords for the search are retrieved. Then the content-matching service conducts the searches within the virtual repository of the ASDI and obtains all of the needed data services for a scientific study. Spatial and temporal subsetting of needed data is also processed by the content-matching service. Nonportrayal services will be converted into the same format as the portrayal services for integration (accomplished by the portrayal service). In addition, the services should be unified with the default projection system, such as EPSG:32633 for Arctic region scientific data (accomplished by the reprojection service). After the unification of all relevant services, the data that are received from multiple services will be integrated by an overlay service for various analysis purposes, e.g., correlation analysis.

4. An ASDI prototype and results

A proof-of-concept prototype (available at http://eie.cos.gmu.edu/VASDI), based on the discussed ASDI architecture and the proposed technologies, was developed to enable the chaining of various scientific data and services. In the current version of the prototype, the GUI browsing using Microsoft IE and Firefox in a Windows Operating System (OS) is supported. When a researcher explores "the influence of solid water dynamics in the Arctic region to bio-habitat," he can start the query by typing in a more intuitive keyword, such as "snow" rather than "solid water." The semantic search service (Fig. 8, Box 1) will generate a chaining workflow to identify all relevant datasets (Fig. 8, Box 2) after the spatial and temporal subset (Fig. 8, Box 4). As shown in Fig. 8, once a query term "snow" is given, knowledge reasoning could infer "Snowfall" as a Synonym; "Precipitation" as a broader term, "Water" as a related substance, "Precipitation," "Cloud," and "Wind" as related Phenomena, "Pressure" as a related Property, and "Deposition" as a related Process. In addition, "Rivers," "Bio_Sample," and "Ice" are automatically inferred for service composition. "Blue Marble" is set as the base map. Through semantic reasoning, scientists have a rich dictionary to choose the best resources they need. The evaluation of the "best" is based on the quality of the discovered data (e.g., response time) and the

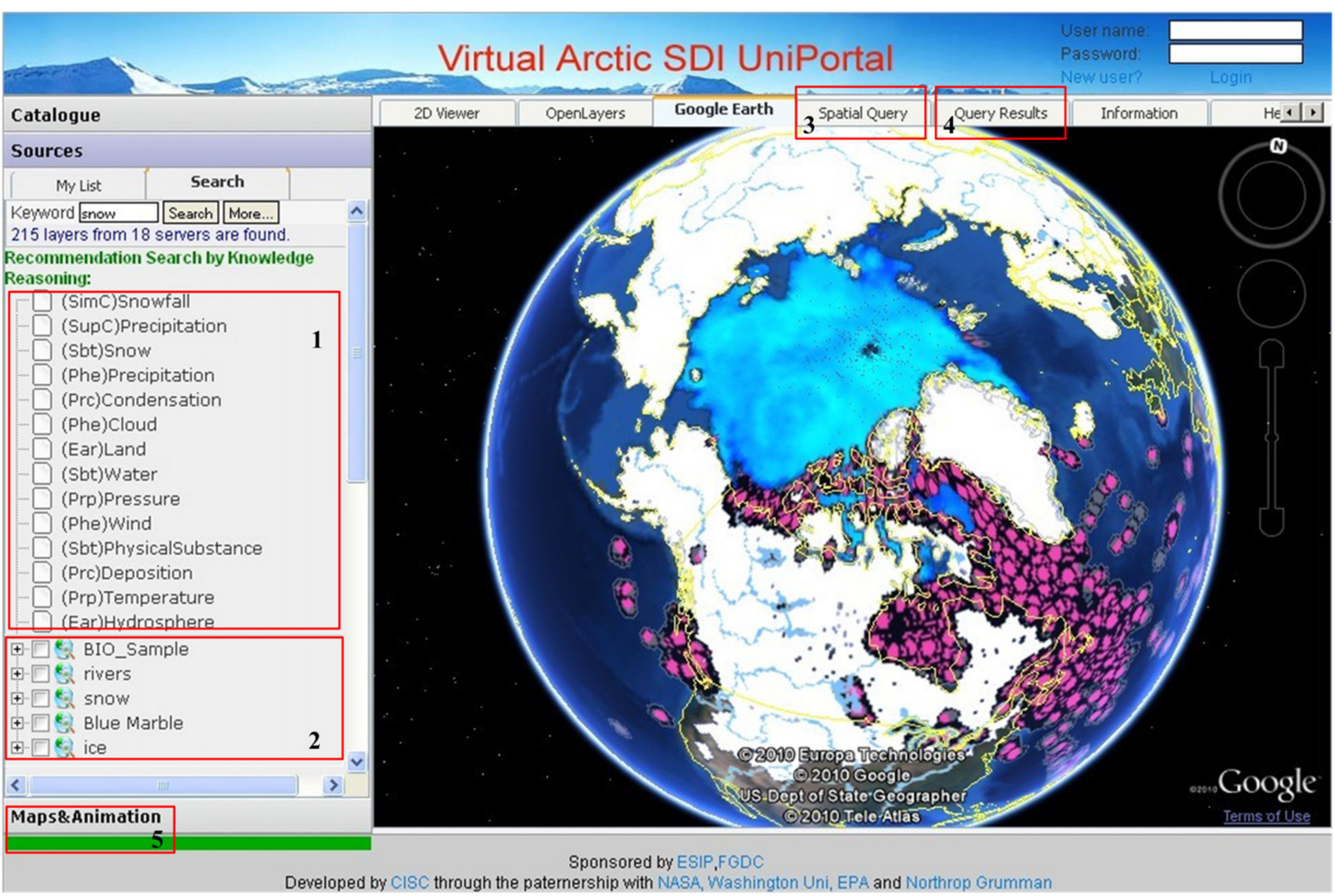

Fig. 8. ASDI prototype. Box 1, semantic query expansion; Box 2, semantic query expansion; Box 3, 3D visualization using Google Earth; Box 4, spatial-temporal subset service; Box 5, animation of time series data.

quality information is indicated by an icon bar displayed on the results panel (Fig. 8, Box 2). After selecting the most suitable service by the client, an integration service is invoked to automatically overlay datasets and display them in a 2-D or 3-D client (Fig. 8, Box 3). The produced imagery in the map client indicates a phenomenon that the bio-habitats are mostly distributed along the coast and within large water bodies. To identify the science principle below that, river data, snow cover, and ice extent can be used. The scientist can also generate a time-series based animation (Fig. 8, Box 5) to discover how the variation of solid water concentration influences the immigration or bio-habitat change.

5. Conclusions and discussion

This paper discussed and addressed three important research challenges (service discovery, knowledge base development, and service decomposition and chaining) when building an integrated ASDI. The proposed hybrid approach, which combines multi-catalogue searching and active crawling mechanisms, helps to collect rich resources to support scientific modeling. The provision of diverse resources from the central access point of our ASDI improves the potential for data sharing. Meanwhile, a hydrology ontology was utilized to provide a controlled vocabulary to improve the effectiveness of the data search and chaining process. The adoption of semantic technology not only improves Arctic scientific data modeling but also modeling of the knowledge contained in the data. It also overcomes the limitation in traditional SDI data discovery and implements some extent of automation to facilitate scientific analysis. The service chaining engine, which combines a centralized control flow pattern and a cascaded control pattern, balances the flexibility in chain node selection and the design complexity. Finally, a proof-of-concept prototype is implemented to demonstrate the capability of the proposed ASDI in facilitating Arctic scientific data discoveries, semantic reasoning, and multiple-dimensional visualization. An example scenario in studying the influence of solid water dynamics to the bio-habitat in the Arctic region highlights the importance of such an infrastructure in improving Arctic studies in a GCI context.

For future research, the area of spatial ranking has attracted our attention. In the proposed ASDI prototype, the problems of collecting and reasoning relevant Arctic resources are well addressed by employing semantic technology. Meanwhile, providing the ‘‘best’’ resources to satisfy the needs of various scientific analyses is also important. In our current implementation, an intuitive performance indicator in terms of availability and response time is used to rank the quality of the datasets retrieved by semantic reasoning. In the future, we will investigate a more comprehensive ranking model, e.g., taking spatial resolution, data precision, support of polar projection, etc., into account, to further improve the Quality of Service (QoS)-based ranking. However, the problem exists in how to collect the QoS indicators from the metadata or the dataset itself. Solving this problem needs the efforts from service providers, the GCI agents, and the data consumers. The service providers should encode more quality information into the metadata when the service is deployed. The GCI agents, such as our proposed ASDI portal, need to implement a tool that can automatically detect the quality information even if they are missing in the metadata description. For example, the agent can detect the spatial resolution of a raster dataset by comparing the amount of pixels per unit of an image map; it can also detect the resolution of a vector datasets by comparing the number of vector points of an object’s shape. Meanwhile, the data consumers will be considered as an important part of quality measurement. In the future, our ASDI portal will provide a scoring system that allows end users to input quality information based on their experience. We will also combine semantic-based relevance ranking algorithms, e.g., algorithms measuring hierarchical and geographic–topological relationship (Göbel and Klein, 2002), into the ranking model to evaluate the semantic relevance between resources retrieved from the ASDI and the users’ need. In addition, we will extend the capability of the proposed ASDI portal to serve the discovery, chaining, and multidimensional visualization (Li et al., this issue) of other scientific data formats/sources, such as OGC Web Feature Service (WFS), Network Common Data Form (NetCDF), and Hierarchical Data Format (HDF).We expect this work to improve the capability of finding better data sources for scientific simulation and prediction.

References

ACIA, 2004. Impact of a Warming Arctic: Arctic Climate Impact Assessment. Cambridge University Press, Cambridge, UK, 139 pp.

Alameh, N., 2002. Service chaining of interoperable geographic information web services. Internet Computing 7 (1), 22–29.

Al-Masri, E., Mahmoud, Q.H., 2007. Interoperability among service registry standards. Internet Computing 11, 74–77.

Barber, D.G., Lukovich, J.V., Keogak, J., Baryluk, S., Fortier, L., Henry, G.H.R., 2008. The changing climate of the arctic. Arctic 61 (Suppl. 1), 7–26.

Barber, D.G., Massom, R.A., 2007. The role of sea ice in Arctic and Antarctic polynya. In: Smith, W.O., Barber, D.G. (Eds.), Polynyas: Windows to the World. Elsevier, Amsterdam.

Bernard, L., Einspanier, U., Haubrock, S., Hübner, S., Kuhn, W., Lessing, R., Lutz, M., Visser, U., 2003. Ontologies for intelligent search and semantic translation in spatial data infrastructure. Photogrammetrie-Fernerkundung-Geoinformation 2003 (6), 451–462.

Bernard, L., Kanellopoulos, I., Annoni, A., Smits, P., 2004. The European geoportal—one step towards the establishment of a European spatial data infrastructure. Computers, Environment and Urban Systems 29 (1), 15–31.

Bowers, S., Lin, K., Ludäscher, B., 2004. On integrating scientific resources through semantic registration. In: Proceedings of 16th International Conference on Scientific and Statistical Database Management. Washington DC, USA, pp. 349–352.

Brickley, D., Guvha, R.V., 2004. RDF Vocabulary Description Language 1.0: RDF Schema. World Wide Web Consortium (accessed 05.05.10.).

Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K., 2004. Jena: implementing the semantic web recommendations. In: Proceedings of the 13th International World Wide Web Conference. New York, NY, USA, pp. 74–83.

Clement, L., Hately, A., von Riegen, C., Rogers, T., 2005. UDDI Version 3.0.2. OASIS UDDI Specification Committee, 420 pp 〈http://uddi.org/pubs/uddi v3.htm〉 (accessed 07.07.10.).

Coleman, D.J., Nebert, D.D., 1998. Building a North American spatial data infrastructure. Cartography and Geographic Information Science 25 (3), 151–160.

Craglia, M., Annoni, A., 2006. INSPIRE: an innovative approach to the development of spatial data infrastructure in Europe. In: Onsrud, H. (Ed.), Research and Theory in Advancing Spatial Data Infrastructure. ESRI Press, Redlands, pp. 93–105.

Dean, M., Schreiber, G., 2004. OWL Web Ontology Language Representation. World Wide Web Consortium. 〈http://www.w3.org/TR/owl-ref〉 (accessed 05.05.10.).

Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A., 1990. Indexing by latent semantic analysis. Journal of the American Society of Information Science 41 (6), 391–407.

de La Beaujardiere, J. (Ed.),, 2004. Web Map Service Implementation Specifications. Version 1.3. OGC Document Number 04-024. OpenGeospatialConsortium Inc., US, 85 pp. 〈http://portal.opengeospatial.org/files/?artifact_id=14416〉 (accessed 05.05.10.).

de Sompel, H.V., Nelson, M., Lagoze, C., Warner, S., 2004. Resource harvesting within the OAI-PMH framework. D-Lib Magazine 10, 12.

Dessai, S., Hulme, M., Lempert, R., Pielke, R., 2009. Do we need better predictions to adapt to a changing climate? Earth Observation System 90, 111–112.

Díaz, L., Granell, C., Gould, M., 2008. Case study: geospatial processing services for web-based hydrological applications. In: Sample, J.T., Shaw, K., Tu, S., Abdelguerfi, M. (Eds.), Geospatial Services and Applications for the Internet. Springer, pp. 31–46.

FGDC, 1998. FGDC Content Standard for Digital Geospatial Metadata v2.0. Federal Geographic Data Committee, 90 pp. 〈http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf〉 (accessed 13.07.10.).

Flato, G.M., Boer, G.J., 2001. Warming asymmetry in climate change simulations. Geophysical Research Letters 28 (1), 195–198.

Fox, P., McGuinness, D., Middleton, D., Cinquini, L., Darnell, J.A., Garcia, J., West, P., Benedict, J., Solomon, S., 2006. Semantically-enabled large-scale science data repositories. In: Proceedings of the International Semantic Web Conference. Athens, GA, US, pp. 792–805.

Friis-Christensen, A., Lucchi, R., Lutz, M., Ostlander, N., 2009. Service chaining architectures for applications implementing distributed geographic information processing. International Journal of Geographic Information Science 23 (5), 561–580.

Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T., 1983. Statistical semantics: analysis of the potential performance of key-word information systems. Bell System Technical Journal 62 (6), 1753–1806.

GEMET, 1999. General European Multi-Lingual Environmental Thesaurus. European Environmental Agency. 〈http://www.eea.eu.int〉 (accessed 05.05.10.).

Georgiadou, Y., Puri, S.K., Sahay, S., 2005. Towards a potential research agenda to guide the implementation of spatial data infrastructures—a case study from India. International Journal of Geographic Information Science 19 (10), 1113–1130.

Göbel, S., Klein, P., 2002. Ranking mechanisms in meta-data information systems for geo-spatial data. In: Proceedings of EOGEO-2002, the 2002 Workshop on Earth Observation and Geo-Spatial Data. Ispra, Italy.

Goodchild, M.K., Egenhofer, M.J., Fegeas, R., Kottman, C. (Eds.), 1999. Interoperating Geographic Information Systems. Springer-Verlag, New York, 536 pp.

Gray, J., Liu, D., Nieto-Santisteban, M., Szalay, A., DeWitt, D., Heber, G., 2005. Scientific data management in the coming decade. ACM SIGMOD 34 (4), 34–41.

Gruber, T., 1993. A translation approach to portable ontology specification. ACMD knowledge acquisition. Special Issue: Current Issues in Knowledge Modeling 6 (2), 199–220.

Hey, T., Trefethen, A.E., 2005. Cyberinfrastructure in e-science. Science 308 (5723), 817–821.

Holland, P., Reichardt, M.E., Nebert, D., Blake, S., Robertson, D., 1999. The global spatial data infrastructure initiative and its relationship to the vision of a digital earth. In: Proceedings of the International Symposium on Digital Earth. Beijing, China, pp. 1–7.

Holzmann, G.J., Pehrson, B., 1994. The Early History of Data Networks 1262 (Perspectives). Wiley-IEEE Computer Society Press, 304 pp.

Huebert, R., 2009. Science, cooperation and conflict in the Arctic region. In: Shadian, J.M., Tennberg, M. (Eds.), Legacies and Change in Polar Sciences: Historical, Legal and Political Reflections on the International Polar Year. Ashgate Publishing, pp. 63–72.

ISO, 2001. ISO 19119:2005 Geographic Information—Services. International Standard Organization. 〈http://www.iso.org/iso/catalogue_detail.htm?csnumber=39890〉 (accessed 05.05.10.).

ISO, 2007. ISO 19139:2007 Geographic Information–Metadata–XML Schema Implementation. International Standard Organization. 〈http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557〉 (accessed 13.07.10.).

Jacoby, S., Smith, J., Ting, L., Williamson, I., 2002. Developing a common spatial data infrastructure between state and local government—an Australian case study. International Journal of Geographic Information Science 16 (4), 305–322.

Johnson, G.W., Gaylordb, A.G., Manley, W., Franco, J.C., Cody, R.P., Brady, J.J., Dover, M., Garcia-Lavigne, D., Score, R., Tweedie, C.E.. Development of the Arctic research mapping application: interoperability challenges and solutions. Computers and Geosciences, this issue. doi:10.1016/j.cageo.2011.04.004.

Klein, D., Manning, C.D., 2003. Fast exact inference with a factored model for natural language parsing. In: Proceedings of Advances in Neural Information Processing Systems 15. MIT Press, Cambridge, MA, pp. 3–10.

Latre, M.A., Lacasta, J., Mojica, E., Nogueras-Iso, J., Zarazaga-Soria, F.J., 2009. An approach to facilitate the integration of hydrological data by means of ontologies and multilingual thesauri. In: Proceedings of 12th AGILE International Conference on Geographic Information Science—Advances in GIScience. Hannover, Germany, pp. 155–171.

Li, W., 2010. Automated Data Discovery, Reasoning and Ranking in Support of Building an Intelligent Geographic Search Engine. Ph.D. Dissertation. George Mason University, 168 pp.

Li, W., Chen, G., Kong, Q., Wang, Z., Qian, C. A VR-ocean system for interactive geospatial analysis and 4D visualization of marine environment around Antarctica. Computers and Geosciences, this issue. doi:10.1016/j.cageo.2011.04.009.

Li, W., Yang, C., Raskin, R., 2008. A semantic enhanced search for spatial web portals. In: Technical Report of Semantic Scientific Knowledge Integration of AAAI Spring Symposium. Merlo Park, CA, pp. 47–50.

Li, W., Yang, C., Yang, C., 2010. An active crawler to discovery geospatial web services—a case study with OGC web map service. International Journal of Geographic Information Science 24 (8), 1127–1147.

Li, Z., Yang, C., Wu, H., Li, W., Li, Miao, in press. An optimized framework for seamlessly integrating OGC web services to support geospatial sciences. International Journal of Geographic Information Science. doi:10.1080/13658816.2010.484811.

Lynch, C.A., 1991. The Z39.50 information retrieval protocol: an overview and status report. ACM SIGCOMM Computer Communication Review 21, 58–70.

Ma, X., Li, G., Xie, K., Shuai, M., 2006. A peer-to-peer approach to geospatial web service discovery. In: Proceedings of First International Conference on Scalable Information System, Article No. 53. Hong Kong, China.

Madsen, J., Tamstorf, M., Klaassen, M., Eide, N., Glahder, C., Riget, F., Nygaard, H., Cottaar, F., 2007. Effect of snow cover on timing and reproduction in high-Arctic pink-footed geese Anser brachyrhynchus. Polar Biology 30, 1363–1372.

Maidment, D.R., 2008. CUAHSI Hydrological Information System: Overview of Version 1.1. Consortium of Universities for the Advancement of Hydrologic Science. 〈http://his.cuahsi.org/documents/HISOverview.pdf〉 (accessed 05.05.10.).

Melbourne, B.A., Hastings, A., 2009. Highly variable spread rates in replicated biological invasions: fundamental limits to predictability. Science 325, 1536–1538.

Nadine, A., 2003. Chaining geographic information web services. IEEE Internet Computing 7 (5), 22–29.

Nauman, M., Khan, S., Amin, M., Hussain, F., 2008. Resolving lexical ambiguities in folksonomy based search systems through common sense and personalization. In: Proceedings of the Workshop on Semantic Search at the Fifth European Semantic Web Conference. Tenerife, Spain, pp. 2–13.

Nebert, D. (Ed.), 2004. Developing Spatial Data Infrastructures: The SDI Cookbook v.2.0. Global Spatial Data Infrastructure (GSDI), Federal Geographic Data Committee, 150 pp. 〈http://www.gsdi.org〉 (accessed 05.05.10.).

Nebert, D., Whiteside, A., 2005. Catalog Services. Version 2. OGC Document Number 07-006r1. OpenGeospatialConsortium Inc. 〈http://portal.opengeospatial.org/files/?artifact_id=20555〉 (accessed 05.05.10.).

Nogueras-Iso, J., Zarazaga-Soria, F.J., Bejar, R., Alvarez, P.J., Muro-Medrano, P.R., 2005. OGC catalog services: a key element for the development of spatial data infrastructures. Computers and Geosciences 31 (2), 199–209.

Notz, D., 2009. The future of ice sheets and sea ice: between reversible retreat and unstoppable loss. Proceedings of National Academy of Sciences of the United States of America 106 (49), 20590–20595.

NSF, 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. NSF Panel Reports, 84 pp.

OASIS, 2002. OASIS/ebXML Registry Information Model v2.0. OASIS Specification Document. 〈http://www.oasis-open.org/committees/regrep/documents/2.0/specs/ebRIM.pdf〉 (accessed 13.07.10.).

Olsen, L., 2001. Supporting interagency collaboration through the global change data and information system (GCDIS). In: Proceedings of the American Meteorological Society. Long Beach, CA, pp. 348–351.

Parent, J., Latour, S., Camfield, A.. Canadian geospatial data infrastructure and Arctic applications. Computers and Geosciences, in this issue.

Prud'hommeaux, E., Seaborne, A., 2005. SPARQL Query Language for RDF. World Wide Consortium. 〈http://www.w3.org/TR/2005/WD-rdf-sparql-query-20050217/〉 (accessed 05.05.10.).

Rajabifard, A., Chan, T.O., Williamson, I.P., 1999. The nature of regional spatial data infrastructures. In: Proceedings, of the 27th Annual Conference of AURISA. Blue Mountains, New South Wales.

Rajabifard, A., Feeney, M.F., Williamson, I.P., 2002. Future directions for SDI development. International Journal of Applied Earth Observation and Geoinformation 4 (2002), 11–22.

Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Computers and Geosciences 31 (9), 1119–1125.

Roesch, A., Wild, M., Gilgen, H., Ohmura, A., 2001. A new snow cover fraction parameterization for the ECHAM4 GCM. Climate Dynamics 17, 933–946.

Singh, D., 2010. The Biological Data Scientist. Business, Bytes, Genes and Molecules. 〈http://mndoci.com/2010/06/22/the-biological-data-scientist/〉 (accessed 15.07.10.).

Singh, G., Bharathi, S., Chervenak, A., Deelman, E., Kesselman, C., Manohar, M., Patil, S., Pearlman, L., 2003. A metadata catalog service for data intensive applications. In: Proceedings of the 2003 ACM/IEEE conference on Supercomputing. Washington, DC, USA, pp. 33–49.

Sorensen, M., Rich, S., Hoover, S., Lawrence, D., 2004. Arctic Spatial Data Infrastructure (SDI) for Scientific Research White Paper. 〈http://www.arcus.org/gis/forumdocs/circumarctic_2004-06-13.pdf〉 (accessed 08.07.10.).

Suranjana, S., Van Den Dool, H.M., 1988. A measure of the practical limit of predictability. Monthly Weather Review 116, 2522–2526.

Tarboton, D., Horsburgh, J., Maidment, D., Jennings, B., 2006. CUAHSI Community Observations Data Model Working Design Specifications Document—V3.1. CUAHSI Org. 〈http://giscenter.isu.edu/training/pdf/geotech_semi〉 (accessed 05.05.10.).

Tisthammer, W.A., 2010. The Nature and Philosophy of Science. UFO Evidence. No. 1386. 〈http://www.ufoevidence.org/documents/doc1386.htm〉 (accessed 08.07.10.).

Vretanos, P.A. (Ed.),, 2005. Web Feature Service Implementation Specifications. Version 1.1.0. OGC Document Number 04–094. Open Geospatial Consortium, 131 pp. 〈http://portal.opengeospatial.org/files/?artifact_id=8339〉 (accessed 05.05.10.).

Wadhams, P., Munk, W., 2004. Ocean freshening, sea level rise, sea ice melting. Geophysical Research Letters 31, L11311. doi:10.1029/2004GL020029.

Weibel, S., Kunze, J., Lagoze, C., Wolf, M., 1998. Dublin Core Metadata for Resource Discovery. RFC 2413. Internet Engineering Task Force. 〈http://dublincore.org/documents/1998/09/dces/〉 (accessed 05.05.10.).

White, D., Hinzman, L., Alessa, L., Cassano, J., Chambers, M., Falkner, K., Francis, J., Gutowski Jr., W.J., Holland, M., Holmes, R.M., Huntington, H., Kane, D., Kliskey, A., Lee, C., McClelland, J., Peterson, B., Rupp, T.S., Straneo, F., Steele, M., Woodgate, R., Yang, D., Yoshikawa, K., Zhang, T., 2007. The arctic freshwater system: changes and impacts. Journal of Geophysical Research 112 (2007), G04S54. doi:10.1029/2006JG000353.

Whiteside, A., Evans, J. (Ed.),, 2006. Web Coverage Service Implementation Specifications. Version 1.1.0. OGC Document Number 06-083r8. OpenGeospatialConsortium Inc., 143 pp. 〈http://portal.opengeospatial.org/files/?artifact_id=27297〉 (accessed 05.05.10.).

Yang, C., Nebert, D., Taylor, F.D.R., Establishing a sustainable and cross-boundary geospatial cyberinfrastructure to enable polar research. Computers and Geosciences, this issue. doi:10.1016/j.cageo.2011.06.009.

Yang, C., Raskin, R., Goodchild, M., Gahegan, M., 2010. Geospatial cyberinfrastructure: past, present and future. Computer, Environment and Urban Systems 34 (4), 264–277.

Yang, P., Evans, J., Cole, M., Alameh, N., Marley, S., Bambacus, M., 2007. The emerging concepts and applications of the spatial web portal. Photogrammetry and Remote Sensing 73 (6), 691–698.

Zhou, C., AI, S., E. D., Wang, Z., Chen, N., Grove mountains recovery and relevant data distribution service. Computers and Geosciences, this issue. doi:10.1016/j.cageo.2011.05.013.