Spatial Data Infrastructures and Linked Data

Carlos Granell
Centre for Interactive Visualization
Universitat Jaume I, Castellón, Spain

Sven Schade
European Commission - Joint Research Centre
Institute for Environment and Sustainability, Ispra, Italy

Gobe Hobona
Centre for Geospatial Science
University of Nottingham, Nottingham, United Kingdom

ABSTRACT

A Spatial Data Infrastructure (SDI) is a type of information infrastructure for enhancing geospatial data sharing and access. At the moment, we face the transition from the service-oriented second generation of SDI to a third generation, characterized by user-centric approaches. This new movement closes the gap between classical SDI and Volunteered Geographic Information (VGI). Public use and acquisition of information provides additional challenges within and beyond the geospatial domain. Linked data has been suggested recently as a possible overall solution. This notion refers to a best practice for exposing, sharing, and connecting resources in the (semantic) web. In this paper, we project the linked data approach to SDI and suggest it as a possibility to combine SDI with VGI. We advocate a Spatial Linked Data Infrastructure, which applies solutions for linked data to classical SDI standards. We detail different implementing strategies, give examples, and argue for benefits, while at the same time trying to outline possible fallbacks. We hope that this contribution will enlighten a way towards a single shared information space.

INTRODUCTION

A Spatial Data Infrastructure (SDI) is a type of information infrastructure for enhancing geospatial data sharing and access. An SDI embraces a set of rules, standards, procedures, guidelines, policies, institutions, data, networks, technology and human resources for enabling and coordinating the management and exchange of geospatial data between stakeholders in the spatial data community (Nebert, 2004; Rajabifard et al., 2006; Masser, 2007). All of these constituent ingredients are not static but dynamic over time. For instance communication and information technologies have dramatically changed since 1990s, when first SDI projects took off (Masser, 1999). SDI is a live, adaptable entity by definition that needs to accommodate periodically to emerging technologies, user relationships, socio-economic contexts and other factors that have influence on SDI developments (Masser, 1999; Rajabijard et al., 2006; Masser, 2007; Budhathoki et al., 2008).

The web itself is evolving from the idea of an open, static repository (Callahan, 1985) to an application platform (Ackland, 2009). Recent advances in web technologies like social networks enable new ways of participation, communication and creativity on the web. It is not surprising then that the Web has changed the way in which we work nowadays. Citizens, experts and non-experts alike, are increasingly participating in the process of generating up-to-date information and collaborating with other in solving-problem tasks. This highlights the matter of a transition, changing role of users, from just mere data consumers to active consumers and producers. Consequently users interact, use and access to information infrastructures in a different way (Ackland, 2009).

The shift in the role of users has been also reflected in the geospatial domain, known as Volunteered Geographic Information (VGI, Goodchild, 2007; Goodchild, 2009). VGI highlights that users are active producers of geographic information rather than passive recipients of geographic information by formal organizations. Budhathoki et al. (2008) has suggested a new SDI generation driven by user needs. However, some recent studies on how social networks are used to sharing data in SDI reveal that users are still poorly connected to data resources of their interest (Omran and van Etten, 2007; van Oort et al., in press). Results also indicate that the position of individuals in a given organization (hierarchy of work relationships) has influences on their potential access to and sharing spatial data, i.e., most relationships are vertical, with few horizontal user-to-user connections between peers.

There is a need then to establish connections among datasets, by linking users with other providers, users with users, and datasets with users and providers. Connecting resources (datasets, metadata, users, providers, organizations, etc.) in such a way would let stakeholders to find out who is actually using a particular dataset. In this vision, linked data has been recently suggested as possible solution for crating such information spaces (Bizer et al., 2009a). The notion of linked data refers to a best practice for exposing, sharing, and connecting data sources in the web[i]. In essence, linked data evolved from research on the semantic web (Berners-Lee et al., 1998) and is concerned with creating links between different data sources, to enable the discovery, navigation, analysis, and knowledge inference of data across disciplinary domains.

Given this context, the motivation of this paper is two-fold:

- The idea of users as data producers is expanding the boundaries of SDI and creates new relationships between the traditional roles of consumer and providers. A network of consumers, producers and providers may benefit and enrich SDI developments, above all at local and regional level where the role of citizens is critical.
- In addition, providing multiple connections between resources helps to make explicit the actual links among stakeholders and at the same time affords alternative search strategies to locate useful content (Gahegan et al., 2009). This potentially alleviates the bottleneck of having just one

> traditional entry point in SDI, as the case of Geoportals (Bernard et al., 2005), and also open up SDI resources to a wider audience. Without suitable provision for these needs, many useful resources will go undiscovered (Gahegan et al., 2009).

Linked data has not yet been regarded in the context of SDI. This paper intends to explore the capabilities of linked data to the geospatial science community, with a special focus on the services in SDI. In a *Spatial Linked Data Infrastructure* resources can potentially connect to the vast repositories of structured geospatial data in linked data space, but also may benefit of establishing relationships not only with data sets but with other types of resources such as metadata, users, providers, organizations, etc. As SDI content expands, though, with user-generated content and linked data sources, it is necessary to consider whether traditional mechanisms are still useful for discovering interlinked resources, understanding these connections, and knowing how to use them appropriately.

We are interested at the question to which degree SDI community and OGC services may co-exist with upcoming semantic web technologies and especially with linked data infrastructures (Bizer et al., 2009a). This would let us answer questions like who is working or connected to a given resource (datasets)? what organizations are serving datasets that cover a certain area extension, city or country? what researchers and users (peers) share my data interests?

The reminder of this paper is structured as follows. Section 2 presents required technical background on SDI and geospatial web services, main projects supporting the linked data initiative, and also the most significant projects aligned to VGI. We include a comparison of the three concepts and outline related work. In section 3, we introduce our projection of the linked data approach to classical SDI and identify a set of challenging issues stemmed from the need of combining both approaches. In section 4 we propose a solution together with three implementing strategies and an according discussion, before we subsequently outline future research directions and wrap-up in the last two sections.

BACKGROUND

This section outlines first background concepts around the underlying topics tackled in this paper –spatial data infrastructure, volunteered geographic information, and linked data—along with a summary table that compare all three fields. Finally, relevant related work is discussed.

Spatial Data Infrastructures

The notion of Spatial Data Infrastructure (SDI) refers to the specialization of information infrastructures for the geospatial sciences (Nebert 2004). President Clinton's executive order to establish a national (US) level SDI (Executive Office of the President 1994) is one of earliest milestones in SDI development. Since then, dozens of SDIs have been developed across the globe, both on national-level and on international-level (Crompvoets and Bregt 2007). INSPIRE (Infrastructure for Spatial Information in the European Community) provides a prominent example on international scale (European Parliament and Council 2007). In total, several billions of dollars are spent on SDI-related activities each year (Onsrud et al. 2004). These activities include technical developments and deployment of applications, but moreover work on institutional arrangements, acquisition and maintenance of geospatial data, and accessibility and usability of geospatial information (Masser 2007).

The origins and objectives of the first generation of SDI are reported by Masser (1999). The first SDI initiatives promoted to make public data available to users (product- or data-driven) and engage potential stakeholders in establishing institutional collaborations and data sharing policies. After the first generation, which last approximately until 2000, the second generation of SDIs is process-oriented

(Rajabijard et al., 2006; Budhathoki et al., 2008). The approach of the second generation is driven by service-based data applications as opposed to the data themselves (Bernard and Craglia, 2005). The role of the SDI user slightly shifted from a passive data consumer to a service receipt (Budhathoki et al., 2008). Nowadays, SDIs describe the notion of service-oriented management, access, and processing of geospatial data; they are implemented using web services (Nebert 2004). Current technical developments and growing interest in social aspects of geospatial information sharing indicate the transition to a third generation of SDI, which is way more user-based, i.e., VGI-oriented (Budhathoki et al., 2008).

In general, web services technology have facilitated data integration and promoted interoperability among heterogeneous distributed information sources. The concept of service, as a self-described, independent (loosely coupled) and standards-based interface unit, takes a central role on current distributed computing paradigms such as cloud computing, grid computing, service-oriented architectures (SOA), cyberinfrastructure and lastly web science and e-infrastructure. The application of e-Infrastructure in scientific research is sometimes referred to as e-Science (in Europe) and e-Research (in Australia and the US). In such platforms applications consume remote, cross-disciplinary and distributed data sources by means of data and processing services.

The geospatial community has been at the forefront of the development of e-Infrastructure for the service-oriented sharing of data and computational capability. SDI exemplify the adoption of the SOA paradigm in the geospatial domain and offer the possibility to access distributed, heterogeneous spatial data through a set of policies, common rules and standards that facilitate interconnecting spatial information users in an interoperable way. Interoperability is a basic requirement for distributed information systems and so it is also critical to Geographic Information Systems (GIS) (Goodchild et al., 1999).

Geospatial web services allow users to access, manage, and process geospatial data in a distributed manner (Zhao et al., 2007). The demand for interoperability has boosted the development of standards and tools to facilitate data transformation and integration, mostly in terms of standard interfaces specified by Open Geospatial Consortium (OGC) and Technical Committee 211 (TC211) of International Organization for Standardization (ISO). Some promising examples of OGC interfaces for geospatial services are Sensor Planning Service (SPS), Web Processing Service (WPS) and Web Coverage Processing Service (WCPS). The Catalogue Services for the Web (CSW) provides the central building-block for data, as well as service discovery and retrieval (OGC 2007b). Linked to the SOA principle of service composition, many standards organizations, industry bodies, and the geospatial research community have paid attention to the effective composition and orchestration of geospatial web services; since geospatial web solutions continue to grow and increase in complexity (Alameh, 2003).

Since 2000 the OGC Interoperability Program has conducted a series of hands-on technical initiatives, called jointly OWS (OGC Web Services), to consolidate the adoption of OGC service specifications by demonstrating the ability of connecting several geospatial web services applied to real-world scenarios. The OWS series has led to the development of new geospatial standards and the refinement of existing ones. In this sense, SDIs provide the infrastructure in which geospatial web services play a facilitating role to wrapping and abstracting data sources and integrating geospatial data and services (Díaz et al., 2009). So far, then, experiments of OGC-based geospatial web services suggests that geospatial data interoperability is possible within the SDI community.

Information retrieval requires data and metadata exchange and processing. Numerous standards for related data (and metadata) and supporting services have been defined on conceptual and implementing level (Nebert, 2004). As SDIs target expert users, data models are complex (for example INSPIRE (2009)) and discovery supports sophisticated querying (OGC, 2005). At the moment, data (and metadata) models are defined in UML and encoded in XML Schema. An abstract structure for data modeling and encoding is provided in form of the Geographic Markup Language (GML, OGC, 2007a). Concrete, domain specific data models are called GML application schema. GML already provides possibilities of including metadata, more sophisticated profiles (e.g. for data and service discovery) are provided separately (OGC 2007b). The two ISO standards 19115 and 19139 provide the most common examples (ISO, 2003; ISO, 2007). A more visualization-centric, less-complex data structure has been standardized in form of KML (formerly Keyhole Markup Language, OGC, 2007c). GeoRSS, a way for adding information about geospatial location to RSS (Really Simple Syndication) feeds, has been acknowledged by OGC, too (OGC 2006).

One example of an SDI initiative that adopts geospatial web service standards is the INSPIRE Directive of the European Commission. During the shaping of INSPIRE, one of the goals of the SDI research agenda was to focus on the importance of SDIs versus other information infrastructures. Context has dramatically changed since then. Not only are IT and computing infrastructures constantly evolving, but most importantly how communication and collaboration among geospatial scientists, researchers, and users in general have changed (Goodchild, 2009). The 2009 PAREDE[ii] white paper for European Data infrastructure —a European effort to promote an e-infrastructure for long-term data preservation and common cross-disciplinary data services—, estimates almost 200 European research and social infrastructures, from bioinformatics, physics and environmental science to social sciences to even YouTube videos. SDI remain still a small community and needs to open its resources (data, services, standards, specifications, etc.) to other bigger communities to tackle multidisciplinary problems that go beyond the boundaries of geospatial information. Disciplines rely on different infrastructures, as for example bioinformatics projects that use grid computing extensively for their distributed processing tasks. In this sense, the geospatial community is already responding to these needs at service level, putting to work together geospatial processing services in grid infrastructures (Lee and Percivall, 2008) (Hobona et al., in press). However, there remains a need for a ‘universal’ approach for indentifying and linking data from different disciplines.

Volunteered Geographic Information

In the previous section, SDIs were described as having evolved over two generations. Crompvoets and Bregt (2007) suggest that data were the key driver for SDI development in the first generation of SDI. However, in the second generation of SDI the use of data and the needs of users are the driving force for the development of SDI. Indeed, the user of geographic information now plays a greater role in the creation of geographic information through enabling technologies such as GPS-enabled mobile phones, geotagging cameras and Web 2.0 applications. The information created through these technologies has led to a concept referred to as Volunteered Geographic Information (VGI) (Goodchild 2007) as a form of user-generated content.

VGI can be defined as geographic information generated through the widespread engagement of large numbers of citizens, each with the ability to add, review and revise the contributed information. One of the motivating factors of VGI is self-promotion (Goodchild, 2007). An example of a service that offers self-promotion is the MyMap feature within Google Maps and the Google 3D Warehouse where users can upload their own generated 2D and 3D maps. Similar to open source software, another motivating factor is the potential economic benefit of a free and open product (Riehle, 2007). An example service that offers clear economic benefits through cost-saving is OpenStreetMap. However, for services such as Wikimapia where contributors are anonymous, an additional benefit could be self-satisfaction. Some examples of services based on volunteered geographic information are:

- *OpenStreetMap* (Haklay and Weber, 2008) is a free interactive map of the whole world, developed and maintained by the general public. Users are able to view and edit the map online. Alternatively, users may upload GPS tracks from their handheld devices. Developers are able to download the complete map for free and integrate it into their own GIS, if they have adequate data storage facilities.

- *Flickr* is a website for sharing pictures and videos online. The website allows users to upload, publish, search and download images and video. The website also offers a web map that enables users to geotag photographs. An indicator of the popularity of geotagging in Flickr, is that over three million photographs were geotagged within the first two weeks of Flickr natively supporting the Geo microformat for geotagging (Suda, 2006). Flickr is provided by Yahoo!
- *Panoramio* is a location-oriented website for sharing geotagged photographs. Some of its key features include a gazetteer that allows users to search by place name. It also offers a web map that presents photographs in their tagged locations. The web map is provided through the Google Maps API. Panoramio is provided by Google.
- *YouTube*, another service offered by Google, allows videos to be also geotagged on upload. This feature enables the videos to be linked to from geo-located pop-ups within Google Earth.
- The *MyMaps* feature offered by Google Maps enables users to digitize points, lines and polygons over aerial imagery and cartographic maps presented by Google Maps. Users are also able to add unstructured attributive information to the digitized geometries, thereby creating geospatial features. The features can be collaboratively edited between multiple users and may also be published for viewing by anyone. The service allows users to export the created data to popular formats such as KML and GeoRSS.
- The *Google Sketchup 3D Warehouse* offers an archiving facility for 3D models created using Google Sketchup. The 3D models can also be imported into Google Earth for viewing in their modelled geo-location. The service also engages volunteers to contribute models towards creating 3D models of cities around the world.
- *Blogs* are online journals that allow maintainers to enter posts and visitors to comment on the posts. The posts are presented in reverse chronological order. Popular blogging services such as Wordpress geotagg posts through addition of geographic coordinates in metadata or through the Geo microformat (Suda 2006). Posts can then be disseminated through GeoRSS.
- *Twitter* is a micro-blogging service that allows maintainers to send posts as text messages from mobile phones. Posts in Twitter terminology are referred to as ‘tweets’. The service offers an API that allows developers to geotag tweets. One of its main attractions is that it allows users to ‘follow’ a particular account. This means that the users are alerted when a new entry is posted.

The geographic information offered by these services is not centrally managed in the same way as traditional geographic information products offered by National Mapping Agencies (NMA) and the like. VGI is maintained through iterative revision by volunteers. This suggests that the accuracy of VGI is not consistent across complete datasets, for example, locations with more contributors receive more frequent updates. This also suggests that the accuracy of VGI is not as high as that of professionally created data. The likelihood of errors in VGI does not appear to be affecting the uptake of VGI, as implied by the ranking of Wikimapia and OpenStreetMap at 1,152 and 12,823 respectively out of 16 million websites crawled by the Alexa information service (http://www.alexa.com) offered by Amazon, Inc. Whether to trust VGI or not, depends on the user and the application to which the VGI is required. For example, to find out information about social events nearby, users would likely adopt VGI. However, to find out where to construct a new road, a user is likely to adopt professionally acquired data. The accuracy of VGI continues to affect the perceived trustworthiness of VGI and has been examined recently (Bishr 2008).

It should be noted that although some of the aforementioned services were developed initially outside the domain of geoinformatics, for example Flickr only adopted geotagging after its launch, it has become evident that location is a key part of the metadata associated with the user generated content. To this end, it is necessary to acknowledge the part played by advances in hardware in making geographically referenced user generated content available. An example of such hardware developments, include the lowering cost of GPS-enabled mobile phones and cameras.

Similarly, advances in web technology have made it possible implement web mapping toolkits. Web mapping APIs, based on Web 2.0 technologies, have made it easier for developers to integrate interactive maps into most websites. Users find it easier to identify locations on a map than to recall the geographic coordinates of a location. Even with such significant developments in VGI services and technologies, there are indications that new innovations are emerging. An example of such innovations is the growing number of location based applications on smart phones. Location-aware applications such as Layar and NearestWiki augment maps and camera previews with VGI (Hill, 2006).

Linked Data

The basic pillars of linked data are essentially traditional web technologies and the usage of light-weight techniques for data model representation. The former resides on the use of Uniform Resource Identifiers (URIs) as reference points. A URI is used to uniquely identify an abstract of physical entity, i.e. a piece of data, as well as a person (Berners-Lee et al., 1998). In this context, the entity is called resource. So-called resolvers can be used to decode a URI to the physical location of the according resource.

Linked data relies on the Resource Description Framework (RDF, Lassila et al., 1999) as basic structure for any form of description. RDF provides means to describe any kind of resource in form of triples (subject-predicate-object) (Klyne and Carroll 2004). A basic typing system for subjects, predicates and objects has been proposed as RDF-Schema (RDF-S) (Brickley and Guha 2004). RDF-S allows for extensions in order to specify domain-dependent subtypes. It provides one way to describe domain vocabularies with its own namespace. RDF is used for implementing the linked data, as a single global model for all data sources.

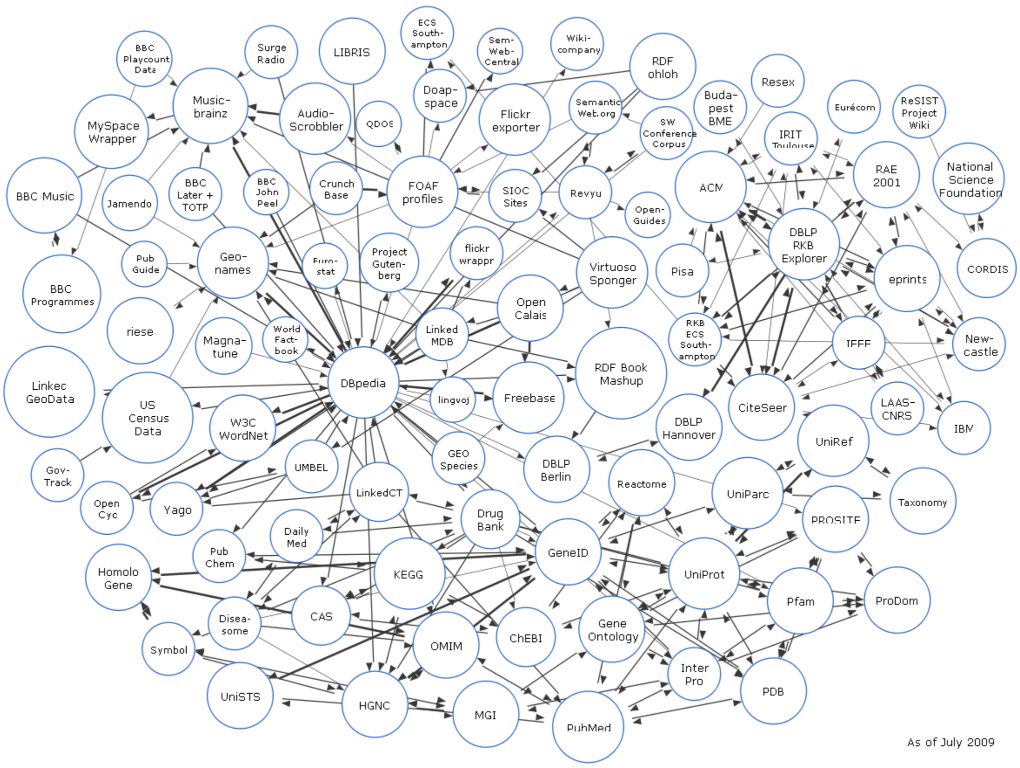

Figure 1. Linked Open Data community project (http://linkeddata.org/).

Figure 1 provides a visual overview of datasets in the linked data cloud and their connections. The size of the circles corresponds to the number of triples in each dataset. The direction of the arrows refers to the dataset that contains the links, e.g., an arrow from A to B means that data source A contains RDF triples that use identifiers from data source B. Bidirectional arrows mean incoming and outcoming links between both data sources.

These datasets are interlinked across several domains. For example, scientific papers and publications are represented by open repositories such as CiteSeer, ACM, DBLP, IEEE, e-prints and so forth. Media content (audio, video, music, etc) comprises a list of data sources like BBC, Musicbrainz, and Audio-Scrobbler. Also, there exists a useful bunch of data sources concerned mainly with the geospatial domain, which the most representative are the following:

- *Geonames*[iii] is a huge gazetteer with nearly eight millions of place names that cover the whole planet. It collects and fusion data from dozens of data sources.
- *Linked Geo Data*[iv] uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the linked data principles. It interlinks this data with other data sources in the linked data cloud (Figure 1).
- *US Census Data*[v] exposes data about population statistics at various geographic levels, from the U.S. as a whole, down through states, counties, sub-counties (roughly, cities and incorporated towns).
- *MONDIAL database*[vi] contains mainly geopolitics data compiled from existing data sources with geographical and global statistics content like the CIA World Factbook (https://www.cia.gov/library/publications/the-world-factbook/) and GeoHive (http://www.xist.org/).
- *Riese*[vii] is a linked data effort to interlinking Eurostat[viii] data sets, European Statistical Information Service that provides a large amount of data on a wide variety of European statistics.
- *British Ordnance Survey*[ix] has recently released the Ordnance Survey linked data project that includes identifiers and names for the administrative and voting areas of UK, and topological relationships between these regions.
- *Telegraphis Linked Open Data*[x] exposes data about countries, capitals and currencies collected from GeoNames and Wikipedia data sources.
- *Linked Railway Data Project*[xi] brings together data on the UK railway network under linked data principles.
- *GEMET*[xii] stands for GEneral Multilingual Environmental Thesaurus. It is the thesaurus recommended by INSPIRE to classify data sets in metadata records, and contains terms related to the environment and environmental data in more than 20 languages.

According to a July 2009 study about the amount of linked data available with each thematic domain (life science, geographic data, publications, media, etc) and the number of established links between data sets (Bizer, 2009), there exists great deal of geospatial data sources in RDF format (nearly 50%) but still poorly connected (scarcely 3%).

Comparison of the Three Notions

In Table 1 we compare SDI, VGI and linked data approaches (columns) with regard to a set of parameters (rows) that fall into the following categories: data model, data access, discovery and publishing. Data model refers to the model used to represent datasets and resources. The second group, data access, is concerned with the strategy adopted to retrieve resources. Discovery comprises the strategy and methods to perform searching. Finally, publishing refers basically to mechanisms to make new resources available.

Each category contains some comparison parameters (rows in Table 1) that basically represent divergences and current open issues for these approaches.

Parameters	**SDI**	**VGI**	**Linked Data**
DATA MODEL			
Rationale	Distributed community data spaces (in possible cooperation).	Great deal of distributed, heterogeneous community data spaces.	A single global data space.
Abstraction unit	Service	Service	Resource
Identification	Service endpoints	Service endpoints	URIs
Representation	• Some but detailed data models (GML, etc.). • Mainly machine readable • Metadata and data separated	• Several but simple data models (Atom, RSS, GeoRSS, geotags, simple annotation, micro-formats etc.). • Human readable • Flat metadata (as tags) and data together.	• RDF model. • Machine readable. • Data and metadata together.
Is it linked?"	Not explicit links.	Not explicit links.	Implicit links.
DATA ACCESS			
Rationale	Metadata access through catalogue services and registries; data access via specialised services.	Data and metadata access via services	Resource access via unique identifier
Mechanism	• Not so simple but standard access. • Detailed service access mechanisms with several parameters • Standard connection bindings: HTTP GET, POST-XML, SOAP.	• Simple but non-standard access. • Multiple APIs depending on concrete service. • Multiple mechanisms depending on service: HTTP GET, POST, SOAP, Restful, proprietary protocol, etc.	• Simple and standard access. • Uniform interface: HTTP GET/POST/PUT/DELE TE • URL dereferencing, just use HTTP and URIs.
Clients	Multiple clients (desktop, web, mash-ups, mobile, etc.).	Multiple services in Web 2.0/Social Web; mash-ups, programmatically.	A few semantic web clients and browsers (not much user-friendly yet).
Is it "linked"?	• No explicit links among SDI sources • Metadata and resource are linked by 'Online Resource' metadata descriptor (if exist). • GML allows for linking elements of the geospatial data model using xlink (rarely used yet).	• No explicit links among VGI sources. • No explicitly. Content is just geo-referenced.	• Resource representation explicitly describes links to related resources.
DISCOVERY			
Rationale	Data and service metadata on centralized repositories.	Searchable resources in the web or via services.	Searchable resources (data) on the very web.
Mechanism	• Spatial queries via OGC Filter against catalogue services. • Simple (keywords, titles) and advanced (complex) search. • Distributed queries and harvesting expand over remote catalogues.	• Diverse discovery methods based often on tag queries. • Simple. • Not cross-source queries, applied to one source at a time.	• SPARQL-based queries, a language for querying RDF databases. • Complex, high expert level required. • Queries expand over several, distributed data sources taking benefit of the RDF graph structure
Clients	Several catalogs clients	Simple interface (tag queries)	Proprietary GUIs for direct

	available.	via services.	SPARQL queries.
Is it "linked"?	• Data and services isolated. Only links between data access services and associated data sets • Common keywords are potentially a means of "linkage".	• Data sets isolated. • Knowledge structures may be derived from folksonomy as a potential means of "linkage".	• Data sources linked. • SPARQL queries exploit in a natural way the notion of link over data sources.
PUBLISHING			
Rationale	Populate metadata on catalogue and exposing (legacy) data via services. Complex tuning.	Uploading, tag and ready.	Transforming (legacy) source data into RDF triples.
Mechanism	• Complex metadata editors and publishing tools (still disconnected). • Expert level required.	• Very simple, intuitive: just tags annotation. • Expert and non-expert alike.	• "RDFising": proprietary solutions on content negotiation translate legacy data to RDF models. • Complex tuning, expert usage.
Clients	Front end to geospatial catalogues, sometimes embedded into metadata editors.	Mainly embedded in web pages with few inputs required.	Clients for establishing links required in the first place.
Is it "linked"?	Not explicit typed links.	Not explicitly typed links.	Publishers have to explicit link their structured data sources with others.

Table 1. Comparison of intensions/capabilities of SDI, VGI, and linked data approaches

In principle, all three approaches are not disjoint but complementary. This means that strengthening connections among them may lead to enriched, valuable, and useful information. However, some open issues and differences for each parameter analyzed in the Table 1 still remain and should be addressed to enable such collaboration. In section Challenges for Spatial Linked Data Infrastructures we discuss in detail some critical issues that fall into the data model category. Concerns like differences in the basic abstraction unit and identification methods as well as links representation and semantics are priority and should be addressed in first term before attempting others regarding access, discovery and publishing. Those latter issues will be treated in section Future Research Directions.

Related Work

The two main areas of relevant related work concern 1) relations between SDI and linked data and 2) techniques and approaches for linking heterogeneous repositories, are overviewed in this section.

In our view that both approaches (SDI and linked data) do not exclude each other, Janowicz et al. (in press) share a similar vision with the notion of micro-SDI, as lightweight linked data applications that still keep the established OGC services for more complex applications. The authors suggest that such a micro-SDI should consist of simplified and lightweight OGC services which can be directly embedded into web pages and applications. Examples towards establishing such a micro-SDI include recent work on next generation gazetteers (Janowicz and Keßler, 2008), a linked data serialization of OpenStreetMap database (Auer et al., 2009), or JavaScript reasoners such as JSExplicit[xiii] which can be directly embedded into web pages to generate context and user-aware information from RDF data on-the-fly.

Hummann-Haidvogel et al. (2009) proposes the usage of multiple coordinated views for searching and navigating web content repositories. The authors collect information regarding a given domain using a web crawler and enrich the local content repository with geospatial, semantic and temporal annotations. Users may visualize annotated content in the local repository via a browser application with multiple coordinated views where any user action (selection, searching, etc.) in one view leads to contextualized updates in the remaining views.

An approach to discovering, describing and understanding resources based on the notion that meaning is carried in the interconnections between resources (data, services, tools, and ontology) and users in the cyberinfrastructure has been provided by Gahegan et al. (2009). Navigation around this universe is achieved by implementing the idea of perspectives as dynamic, conceptual views defined by SPARQL-like queries against a knowledge collection. The authors describe a means to represent a wide variety of interactions between resources using the notion of a knowledge nexus.

Becker and Bizer (2009) have recently proposed DBPedia Mobile, a location –aware semantic web client for mobile devices. This application makes use of geo-related data sets that have been published and interlinked in the context of the linked data cloud.

CHALLENGES FOR SPATIAL LINKED DATA INFRASTRUCTURES

This section begins with a use case for a Spatial Linked Data Infrastructure. Afterwards, we identify related challenges and potentials in connecting geospatial web services and SDI content with emerging linked data and VGI communities. Table 1 highlighted some divergences among these approaches grouped in four categories. In this section, we look especially into the first category, the data model to focus our analysis mainly on the concerns of how to establish links among data sources that have different data models.

The current impediments and challenges for data model discussed in this section are (1) how to deal with the problem of identifying resources in these three approaches? (2) How to represent and where to place links among data sources? (3) How to interpret the semantic of these links?

Use Case

To illustrate the potentials of SDI, VGI and linked data working in the same space and how this approach could potentially help in opening up SDI to a broader non-expert community like VGI and semantic linked data sources, we examine the following scenario as a running example throughout the paper.

A user is planning a short trip to the city of Nottingham (UK) and wishes to get as much as possible informed about this city. She may start off using a semantic web search engine like the Falcons[xiv] search engine and simply types “Nottingham” in the search text box, as many users do with Google search engine. The Falcon engine provides related information about this term (concept) like the corresponding resource for the city of Nottingham in DBpedia[xv] (http://dbpedia.org/resource/Nottingham). By following the link, she is prompted with a great amount of RDF-based data about the city of Nottingham rendered in HTML format. Rather than scrolling down a long HTML page, she may use some linked data browsers available like OpenLink Data Explorer[xvi] that helps users in navigating and inspecting structured content, as illustrated in Figure 2.

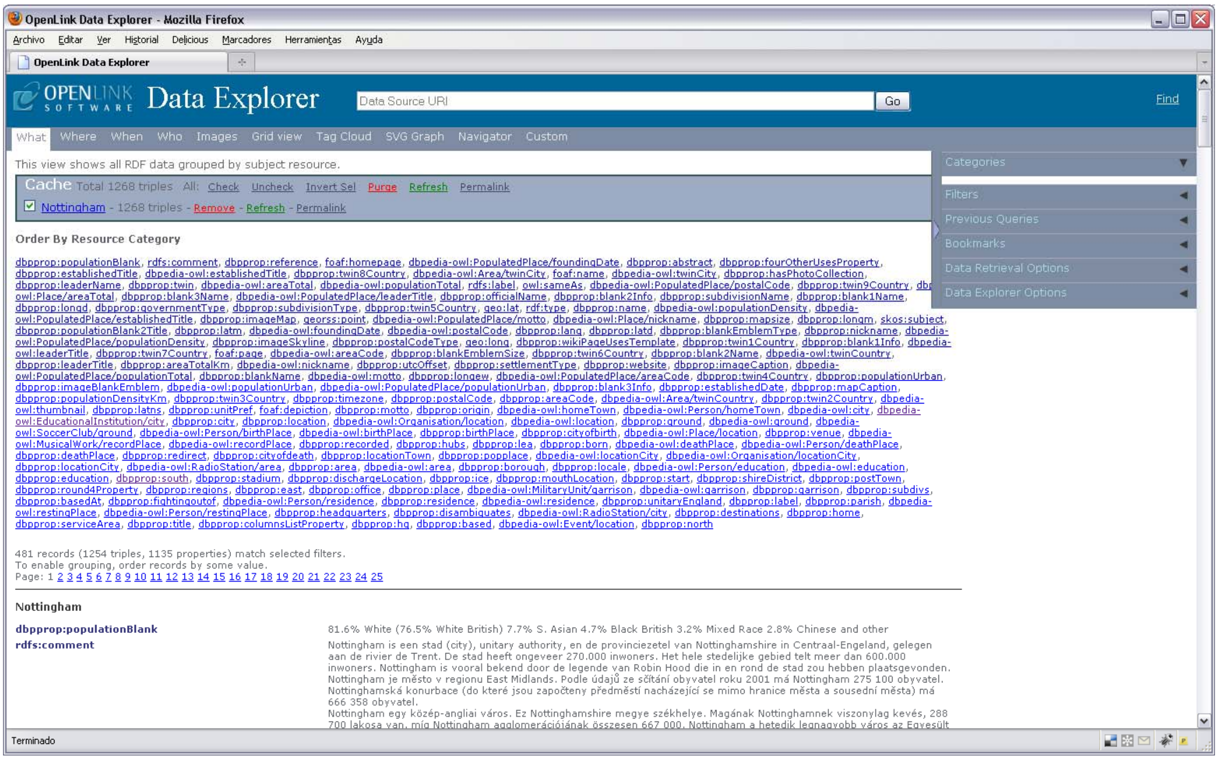

Figure 2. Linked data browser displaying structured content of the resource Nottingham in DBpedia.

DBpedia data source describes almost 3 million of things including any sort of resources like people, places and organizations (Bizer et al., 2009b). Well-known resources like the city of Nottingham may contain a great deal of RDF triples, as shown in Figure 2. Some of these properties contain an implicit geographic connotation like *dbpedia_owl:twincity*, *dbpedia_owl:coordinates*, *dbproop:subdivisionName*, and *dpprob:location*. Connections to the same resource between different data sources are also available mostly by means of the *owl:sameAs* predicate, which indicates that two features actually refer to the same real-world entity.

From the current resource in DBpedia, she may leap to other data sources in a transparent way just by following the typed links. Figure 3 illustrates some existing connections between the Nottingham resource in DBpedia and other linked data sources. In particular, we focus here on data sources that have to do with geospatial contents like places, locations, regions, boundaries, etc. For instance, one of the *owl:sameAs* typed links brings her to Geonames data source. The counterpart Geonames resource for Nottingham city is reachable by dereferencing the (target resource) object property http://sws.geonames.org/26411170 for the predicate *owl:sameAs*. Note that from this point, it is also possible both to navigate to other services linked to the current Geonames resource as for instance a Google map viewer centered on the city and to access to other geo relationships belonging to the Geonames ontology[xvii] like administrative hierarchy for countries (*geoNames:parentFeature*), neighbouring countries (*geoName:neighbour*) and nearby features (*geoName:nearby*). Accessing to other related linked data sources like Linked Geo Data and Ordnance Survey happens on the same basis. Relationships built upon VGI vocabularies like the Friend of a Friend (FOAF[xviii]) are a first attempt to connect linked data and VGI. Relationships like *foaf:based_near* and *foaf: depiction* permit links but still in a limited, punctual sense.

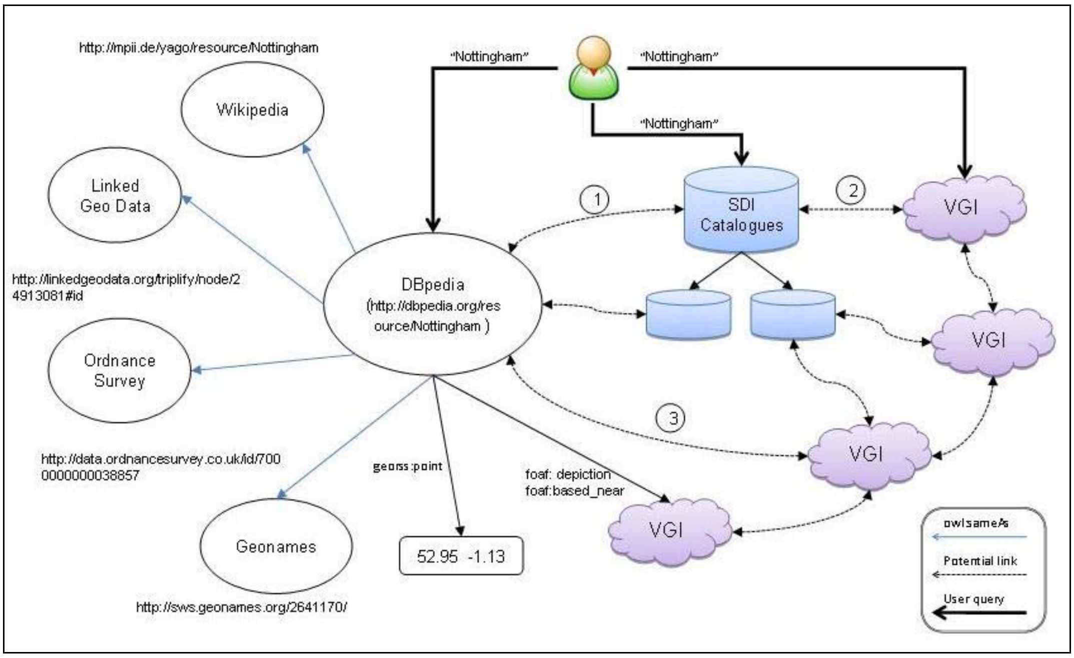

Figure 3.SDI, VGI and linked data sources interrelated around the resource "Nottingham".

Alternatively, she may choose another searching path to discover data of her interest by using the SDI catalogue services. Through a catalogue service, she is able to locate thematic data sets, ortofotos and SDI services that cover the area of Nottingham. In addition, she may turn into social networks and VGIO service to search for Flickr images tagged with the term "Nottingham" (http://www.flickr.com/search/?q=Nottingham).

This example scenario shows that geospatial structured data is available to be crawled by eager users. Why then the DBpedia resource of city of Nottingham contain no links (dotted line numbered with 1 in Figure 3) to data catalogues with metadata about such area, or web map services that serve maps with Nottingham area contain in this bounding box, or thematic layers about related form WFS in, or at least a list of SDI services located in UK. And vice versa, suppose now the same user querying a metadata catalogue with Nottingham as keyword. Resulting metadata records would contain "linked data" descriptors with typed links to other both catalogues and external structured data sources.

It seems obvious that linked structured data has lots of benefits of a wide range of users. Users may choose DBpedia as entry point and go through SDI catalogues to find WFS services of their interest. Or, maybe, users accessing to Geoportal may reach Geonames data sources just following the links. Or just scanning the Flickr feeds (search results) would give enough clues to know about SDI services covering the area depicted in the picture. In essence, alternative search paths would enhance the discovery, access, and browsing of SDI data and services to a border community to exploit_conveniently SDI, VGI and linked data to enrich geospatial applications. In the following we analyze arising challenges derived from the proposed user case.

Challenge 1: Resource Identification

This section examines the issue on how to address resources univocally. In the linked data context, a connection takes place between a source resource and a target resource, as shown in Figure 4. A source resource, like the Nottingham city resource in the Geonames data space (S1), is referenced by a unique

URI (http://sws.geonames.org/2641170). A target resource, like the same resource in DBpedia (T) data source, has also its own URL (http://dbpedia.org/resource/Nottingham). As both resources are identified properly, it is straightforward to connect them through a typed relationship (C1), like *owl:sameAs*. Similarly, the Nottingham resource (S2) in the Linked Geo Data space also may link to the same target resources (T). Creating links is a matter of identifying unambiguously the both edges of the link, the source (subject) and the target (object) resource.

Figure 4. Connections between addressable resources.

Setting a connection then requires first to respond a couple of questions: what is considered a resource itself (abstraction unit parameter in Table 1) and how to identify it (identification parameter in Table 1)? As aforementioned, from the linked data viewpoint, the first question is straightforward addressed since a resource can be anything that it is worth sharing with the community. For instance, looking at the DBPedia space one can encounter movies, films, actors, songs, musicians, cities, institutions, individuals, football teams, and so forth. Indeed, the web itself functions in this way since it contains a great deal of heterogeneous resources (web page, doc, image, video, etc). Every single web resource is univocally identified by its URI. Linked data also relies on the URI mechanisms to identify any resource. Every single resource that might be of interest for the community should be identified unambiguously so that a user using its URI is able to access it. These simple tenets drive the foundations of linked data sources and support the existence of connections.

In contrast to linked data, in the SDI context the basic abstraction unit is the notion of service. Geospatial datasets are discovered, accessed, and shared though services. In this sense, the exercise of comparing linked data and SDI is similar to compare service-oriented and resource-oriented architectural styles. The former is based on the service abstraction, while the latter is rooted by the resource abstraction. SDI provides a single, directly accessible service abstraction that often represents multiple different potential resources from the resource-oriented viewpoint. For instance, features offered by a WFS service might be regarded candidate resources from a resource-oriented perspective.

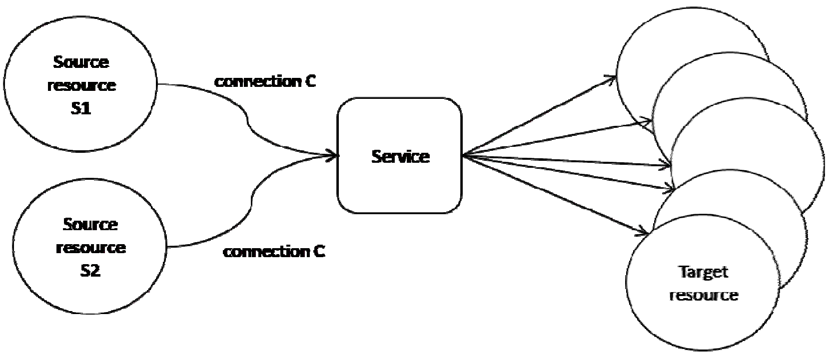

Figure 5. Connections in SDI

SDIs do not provide an explicit, univocally resource identification as linked data does. Figure 5 shows how a source resource establishes a connection with a SDI service, addressable with a URI that refers to a service endpoint. Conceptually, the service acts as a proxy, receiving all incoming connections and bypassing access to the internal data sets. In this scenario, the target resource has always the same URL of the service point, which hides and centrally manages various resources through service operations like *getCapabilities*, *getRecords*, and *describeProcess*. In practice, all target data sets in backend are hidden to external entities and thus connections can be only established between source resources and the proxy service. This suggests that direct connections between resources within linked data and data sets in SDI cannot be set.

The divergence on the basic unit —resource abstraction versus service abstraction— and on the meaning of the URI in each case remains challenging issues. First it is necessary to determine clearly what is considered a resource in the SDI landscape: feature, data set, a collection of data set, metadata files, a collection of metadata files, an image file, etc. Second, every single resource should be addressed unambiguously in such a way that referencing its URL lets access to the very resource. Adopting a resource-oriented architecture would helps to add context and intelligibility to the SDI content (Lucchi et al., 2008). An alternative solution to overcome this issue would be to augment SDI services with URIs resolvers, capable of decoding and creating valid URIs at request-time to the data resources they offer.

Challenge 2: Link Representation

This section examines the issue of where to place a link representation. Resources may have multiple representations. When dereferencing a non-informational resource (real-world entities like person, city, etc.) one does not get the resource itself, but a representation of the resource as a description in a certain format. In the running scenario of Figure 3, the action of dereferencing the resource identifier http://dbpedia.org/resource/Nottingham in a browser renders the resource description in RDF/XML format. Such a description contains links to other resources.

The trend in SDI is to describe data and services through the use of metadata. So, metadata descriptions are *a priori* the best place where links representations and pointers to external resources should be. This implies a need to analyze the capabilities of metadata descriptions in SDI to contain pointers to other SDI resources like metadata records, datasets and services. In the following we deal with the questions on how metadata records contain links to datasets they describe and/or external structured resources.

Do data metadata records contain links to other SDI resources?

Metadata descriptions are stored according to the international standard for geographic information metadata ISO19115 (ISO 2003). Metadata based on ISO19115 can be presented in a variety of human readable formats such as HTML and plain text. To support the structured encoding of metadata, ISO19139 (ISO 2007) was developed to offer an XML schema for metadata based on ISO19115. The XML schema offered by ISO19139 makes it easier for applications to identify metadata elements because it formalizes the names and structure of the metadata elements. Where appropriate, restricted values for metadata elements are also enforced as 'enumerations' for example the *topicCategory* field in ISO19115. Whereas, the historic use of metadata has been to describe data, the uptake of services as key resources within SDI has led to the development of an international standard for geographic information services, ISO19119 (ISO 2005), which includes metadata for describing services. It should be noted that both the ISO19119 and the ISO19139 reference the ISO19115 as the base standard for metadata. Therefore, unless otherwise stated, the metadata elements described in the rest of this section are found in ISO19115.

Dataset title (M) (MD_Metadata>MD_Identification.citation>CI_Citation.title)	Spatial representation type (O) (MD_Metadata>MD_DataIdentification.spatialRepresentation)
Dataset reference date (M) (MD_Metadata>MD_Identification.citation>CI_Citation.date)	Reference system (O) (MD_Metadata>MD_ReferenceSystem)
Dataset responsible party (O)	Lineage (O)

(MD_Metadata>MD_Identification.pointOfContact>CI_ResponsibleParty)	(MD_Metadata>DQ_DataQuality.lineage>LI_Lineage)
Geographic location of the dataset (by four coordinates or by geographic identifier) (O) (MD_Metadata>MD_DataIdentification.extent>EX_Extent>EX_GeographicExtent>EX_GeographicBoudingBox or EX_GeographicDescription)	On-line resource (O) (MD_Metadata>MD_Distribution>MD_DigitalTransferOption.online>CI_OnlineResource)
Dataset language (M) (MD_Metadata>MD_DataIdentification.language)	Metadata file identifier (O) (MD_Metadata.fileIdentifier)
Dataset character set (C) (MD_Metadata>MD_DataIdentification.characterSet)	Metadata standard name (O) (MD_Metadata.metadataStandardName)
Dataset topic category (M) (MD_Metadata>MD_DataIdentification.topicCategory)	Metadata standard version (O) (MD_Metadata.metadataStandardVersion)
Spatial resolution of the dataset (O) (MD_Metadata>MD_DataIdentification.spatialResolution>MD_Resolution.equivalentScale or MD_Resolution.distance)	Metadata language (C) (MD_Metadata.language)
Abstract describing the dataset (M) (MD_Metadata>MD_Identification.abstract)	Metadata character set (C) (MD_Metadata.characterSet)
Distribution format (O) (MD_Metadata>MD_Distribution.MD_Format.name and MD_Format.version)	Metadata point of contact (M) (MD_Metadata.contact>CI_ResponsibleParty)
Additional extent information for the dataset (vertical and temporal) (O) (MD_Metadata>MD_DataIdentification.extent>EX_Extent>EX_>TemporalExtent or EXVerticalExtent)	Metadata data stamp (M) (MD_Metadata.dateStamp)

Table 2. Core metadata ISO 19115

In the ISO19115 metadata model hundreds of fields are offered. However, as acknowledged in the standard “typically only a subset of the full number of elements is used” (ISO 19115:2003, pp. 15). The standard there identifies a minimum set of metadata elements that should be maintained for a dataset. This minimum set of metadata elements are referred to as the core metadata for geographic datasets are presented in Table 2. The table presents the name of each core metadata element in bold and identifies whether that particular element is mandatory (M), optional (O) or Conditional (C). The hierarchical structure of the parent packages of the elements are also presented in descending order with a ‘>’ sign, for example, the temporal extent is part of the extent package, which is part of the *dataIdentification* package and so on. The two-letter abbreviations before the name of the element identify which package that element belongs in, for example *CI* is the *Citation* package and *MD* is the *Metadata* package.

ISO19115 contemplates some metadata descriptors to provide the context necessary to reference a metadata (catalogue) record to the corresponding data or service:

- Filling the *MD_Distribution->transferOptions->MD_DigitalTransferOptions->onLine->CI_OnlineResource* definition.
- Or setting the *MD_Metadata.dataSetURI* attribute value (not listed in Table 2).
- Or setting the *MD_Metadata.fileIdentifier* and *MD_Metadata.parentIdentifier* attribute values.

The first option for linking to a resource is the *CI_OnlineResource* class defined in ISO19115 as the element in charge of containing information about online sources from which the dataset, specification, or community profile name and extended metadata elements can be obtained. The *CI_OnlineResource* class permits to augment a URL in the *linkage* mandatory element with optional values for service definition in *protocol* and *applicationProfile* elements, providing thus a way to link data metadata records to the service. The values contained in this metadata descriptor provide the link to associated datasets in terms of query parameters to the appropriate service. This descriptor is crucial to avoid broken links among SDI resources. The *CI_OnlineResource* metadata descriptor has the following inner elements:

- *linkage* (mandatory) points to the URL from which the dataset or service endpoint can be accessed. In some instances, the linkage may point to the *GetCapabilities* document of a service, which itself presents the service endpoint and additional information.
- *name* (optional) presents the title of the resource (Coverage, FeatureType, Observation, etc.) in free text.
- *protocol* identifies the network binding for connecting to the resource (e.g. HTTP), specified in free text.
- *applicationProfile* (optional), describes a name of an application profile that can be used with the online resource, also in free text.
- *description* (optional) of what the online resource does, also in free text.
- *function* (optional) is a controlled list of the online resource's actions.

Apart from the linkage and function attributes of the *CI_OnlineResource* class, all other attributes are defined as free text, meaning that the values are not from a controlled vocabulary or format. Whereas 'free text' values ensure flexibility in the types of values that can be assigned to a metadata element, applications are not able to trace the resources to which 'free text' values point. The following example is taken from ISO19139 as demonstrating how to reference a dataset online:

```
<onlineResource>
 <CI_OnlineResource id="ID00004">
  <linkage>
   <URL>http://geoengine.nga.mil</URL>
  </linkage>
  <protocol>
   <gco:CharacterString>http</gco:CharacterString>
  </protocol>
 </CI_OnlineResource>
</onlineResource>
```

For services it is not enough to offer the URL of the service endpoint. It is necessary to also provide a description of the standard and version supported by the service, for example, version 1.0.0 of the Web Feature Service offers fewer capabilities than version 1.1.0 of the same standard. The following example is taken from the OGC Web Feature Service specification as demonstrating how to describe a service in ISO19119 metadata presented in the *GetCapabilities* document of a service:

```
<ows:ServiceIdentification>
 <ows:Title>WFS</ows:Title>
  <ows:ServiceType>WFS</ows:ServiceType>
   <ows:ServiceTypeVersion>1.1.0</ows:ServiceTypeVersion>
   <ows:ServiceIdentification>
…
    <ows:Operation name="GetFeature">
    <ows:DCP>
    <ows:HTTP>
    <ows:Get xlink:href="http://www.BlueOx.org/wfs/wfs.cgi?"/>
    <ows:Post xlink:href="http://www.BlueOx.org/wfs/wfs.cgi"/>
   </ows:HTTP>
  </ows:DCP>
…
</ows:Operation>
```

The preceding examples of metadata suggest that for the *CI_OnlineResource* class to provide enough information to bind to a service, it is necessary to include information about the service name, type and version in the protocol attribute of the class. This would allow a client application to bind to a service, if the client supports the specified service type and version. An approach for specifying the service type and version in the protocol attribute of the *CI_OnlineResource* class using a Uniform Resource Name (URN) is as follows:

```
<onlineResource>
 <CI_OnlineResource id="ID00001">
  <name>WFS</name>
  <linkage>
   <URL>
http://www.BlueOx.org/wfs/wfs.cgi?service=WFS&request=GetCapabilities</URL>
  </linkage>
  <protocol>
   <gco:CharacterString>urn:ogc:serviceType:WebFeatureService:1.1.0
   </gco:CharacterString>
  </protocol>
 </CI_OnlineResource>
</onlineResource>
```

URN refers to the subset of Uniform Resource Identifier (URI) “required to remain globally unique and persistent even when the resource ceases to exist or becomes unavailable.” (Berners-Lee et al., 1998). Therefore, a URN does not need to be linked to an existing resource online. An example of the application of URN in SDI is presented by the OGC Best Practice document on the ‘Definition identifier URNs in OGC namespace’[xix]. For linking to resources online, the *linkage* attribute of the *CI_OnlineResource* class adopts URL as values. URL are a subset of URI “that identify resources via a representation of their primary access mechanism (e.g., their network “location”), rather than identifying the resource by name or by some other attribute(s) of that resource.” (Berners-Lee et al., 1998). An application can automatically bind to a service if both the linkage and protocol attributes of the *CI_OnlineResource* class are filled with URL and URN values respectively. However, if ‘free text’ is used for the protocol, then a human is required to interpret the metadata and build an HTTP query to the concrete service that‘s serving the actual dataset.

As the *CI_OnlineResource* entity is optional it may be empty and then there is no way from the user perspective to locate the datasets described by the associated metadata record. The current trend is that users obtain the requested core dataset metadata record from the service catalogue but often fail to visualize the actual dataset because the *CI_OnlineResource* metadata descriptor is missing or badly documented. Users are often prompted with an alert window that current dataset is not accessible.

Another option for referencing resources in metadata is through the optional *dataSetURI* attribute of the main *MD_Metadata* class. The URI value of an *MD_Metadata.dataSetURI* attribute is typically a URL (service endpoint) in order to link to the service. This simple but direct approach may hold linking representations in a variety of formats (RSS, RDF/XML) since, according to the ISO 19115 specification, the data type of the attribute is free text. It is however, necessary to highlight that a URL would need to be augmented with a description of the protocol needed to bind to the described service. In this group also fall the ISO 19115 *MD_Metadata.fileIdentifier* attribute for identifying the particular metadata record and the *MD_Metadata.parentIdentifier* attribute for identifying the metadata record from which the source metadata is a subset. Filling both the *MD_Metadata.parentIdentifier* and *MD_Metadata.fileIdentifier* attributes enables a user to link metadata records that are related through a set-subset relationship.

The early generations of SDI placed particular significance on documenting metadata for datasets. The evolution of SDI and the key role played by geospatial web services suggests that the *CI_OnlineResource* metadata element, should no longer be 'optional'. Instead, if a dataset is not available through a service but through physical transfer (e.g. by courier), a specific URI could be associated with that transfer mechanism (e.g. urn:royalmail:parcelforce:sameday). We note that the version 3 of the INSPIRE implementation rules for metadata specify that the resource locator is "mandatory if linkage to the service is available" (page 14). It is therefore necessary for metadata creation tools to enforce the filling of the *CI_OnlineResource* automatically or to support the service provider in doing so.

High quality spatial data is required to address current issues concerning geospatial information like environment sustainability, climate change and disaster management. The ability to find and access appropriate information within SDIs relies basically on having up-to-date metadata. Rajabifard et al. (2009) have suggested the use of integrated metadata directories to automatically generate metadata that would be maintained also dynamically. In this case, metadata descriptions would mirror continuously the evolution of spatial data sets and keep track of service link updates.

Do data metadata records contain links to external resources?

Looking at the core metadata description (see Table 2), there are a lot of potential links that could be established to improve connectiveness to external data resources, such as the people responsible for datasets or keywords in a common thesaurus. Let's examine carefully some descriptors.

The *CI_ResponsibleParty* class belonging to the *MD_Identification.pointOfContact* entity contains the identification of people and institutions associated with the resource. The *CI_Contact* class offers attributes for the phone number, address, online resource, hours of service and contact instructions for the person or organization responsible for a resource. The *onlineResource* attribute of the *CI_Contact* class accepts values based on the *CI_OnlineResource* class, therefore, the attribute allows for referencing by URL. A first attempt is that these metadata descriptions may be related to people and institutions using the FOAF vocabulary, which is encoded in RDF. Primary predicates like *foaf:Person* and *foaf:organization* may be used to designate relationships to physical people and institutions and the corresponding dataset metadata record. Other secondary predicates can also be used as for example *foaf:homepage*, *foaf:name*, *foaf:depiction*, and *foaf:page*, to provide more connections. By virtue of supporting URL values, persons represented in FOAF can be referenced in the linkage attribute of the *CI_OnlineResouce* class.

However, not only FOAF data sources are suitable. Attribute values within the *CI_Address* class entity such as *city*, *country*, *postalCode* and *administrativeArea* are good candidate to be linked to external resources in target data sources like Dbpedia, Geonames and Linked Geo Data.

INSPIRE recommends the use of the GEMET taxonomy for categorizing metadata records through the *topicCategory* attribute in *MD_DataIdentification* entity class. Indeed, the GEMET terms fit nicely with data sources in life science domain (see Figure 1). Similarly, the *MD_Keywords.keyword* attribute can be used as a "linkage tag" to connect related items in VGI services like Flickr and Panoramio. Linkage can be enhanced by using keywords from a commonly used thesaurus such as GEMET or SWEET (Raskin, 2005). SWEET is a unilingual ontology offering concepts and properties of concepts. The *type* and *thesaurusName* attributes of the *MD_Keyword* class allows for the referencing of concepts in GEMET or SWEET.

In addition to the aforementioned thematic metadata elements, the geographic extent of the dataset described by the metadata is specified in the *EX_GeographicDescription* element. The extent can be expressed as an identifier in the *geographicIdentifier* attribute of the *EX_GeographicDescription* class. The geographic identifier is defined as 'free text' therefore a URI could be used to reference a location.

The challenge for linking metadata to external resources is to increase the set of metadata elements that adopt URI whilst reducing those that use 'free text'. This will make metadata more machine readable, thereby enabling for links to be made from user needs, to services, to data and then to geographic features contained within the data. VGI services, such as Flickr and YouTube, already maintain persistent identifiers for resources they offer. Therefore, metadata creation tools should be enabled to support the referencing of resources offered by VGI where appropriate.

A further challenge for metadata publishers however, will be to ensure that the references remain valid even though the external resources are maintained by private citizens. For example, a metadata record of satellite imagery of a natural flood could be extended to reference Flickr photographs taken by flood victims at ground level. Whereas the metadata record is typically maintained by professionals, there is no guarantee that the Flickr photographs referenced by the metadata record will remain available. Therefore, it will be a challenge for SDI to adapt changes in references to resources in VGI. However, the significant amount of information offered by VGI makes this challenge worthwhile.

Challenge 3: Link Semantic and Context

Describing the model used to encode the range of data links is a general concern of the linked data paradigm. Linking data does not specify the meaning of elements, which are pointed to. Connections make only sense if linking to data in cases where the intended interpretation is well known. For example, a text may contain the term “Nottingham”, which is subject to a set of links (see also Figure 3). One linked may be termed “description” and points to a Wikipedia article about the city. Another predicate uses the GeoRSS vocabulary and as object the according point. Another may use the predicate *os:feature* to point to some resource with the id http://data.ordnancesurvey.co.uk/id/7000000000038857, a fourth link may use a vocabulary called “MyApplicationSpecificSchema”, which points to an instance of a very complex data model for storing the transport network of the city of Nottingham. It depends on the information consumer, which of these likes can be processed in a meaningful way.

With RDF, as well as GML, the structures used to represent the flexible content have to be communicated. This may either happen in a pre-deployment phase, i.e. the meaning of used data structures is fixed and well known before accessing the resource, or at run-time. Well known meaning can be defined as part of user requirements, as for example in the domain of reporting, or by detailed specifications using the context of the used links. One example for the reporting issues for geospatial science considers the environment in Europe. It is known as SEIS (Shared Environmental Information System). Here, required content is pre-defined in form of legal documents and obligations. Otherwise, in the case of RDF, the range of a link can be defined according the types provided in the vocabulary (defined in RDF-S). Yet, the type system has to be documented outside the given model. The documentation of GeoRSS provides one example (http://www.georss.org/Main_Page): “A point contains a single latitude-longitude pair...” The ambiguity in interpretation and the need for human intervention remains challenging. GeoRSS does not specify what the point represents. We do not know if the point associated with Nottingham represents the centroid of the administrative area, the location of the town hall, or any arbitrary point. Adding ontologies, i.e. rich vocabularies, which allow for logical specification of intended interpretations of terms, addresses this issue only to some extent (Guarino 1998). While the set of possible interpretations becomes reduced, the problem of defining intended interpretations of ontological primitives remains (Harnad 1990).

Considering GML and its capabilities of using xlink, as alternative to RDF, such typing possibilities are not given by the current specifications. Here, the type of used GML property may define the data model for the range of the link. For example, if a link is embedded in the point property, the referred URI should follow the definition of a GML point data type. Such an approach requires validation mechanisms and powerful data parsing capabilities. Enabling run-time exploitation of data structures (or models) is even more challenging. It requires inclusion of data model information inside the link itself or to the range of the link. Solutions may include extended use of MIME times or explicit additional links to data models. Re-visiting the example from above, the subject "Nottingham" may be linked via the predicate *myApplicationSpecificSchema:transportNetwork* to a object of type *XML:GML: MyApplicationSpecific-Schema*. This indicates that the range of the link is an *XML* element, following the general rules of *GML* and a more specialized encoding, which is defined as *MyApplicationSpecificSchema*. Depending on client capabilities the link range can be processes, i.e. parsed and analyzed, to certain extent. The use of MIME types implies a need of management and introduction of types per application specific data model. Obviously this solution faces a large overhead. Using links (annotations) to the used model is another approach, which still has to be explored (Maué et al., 2009).

Overall, data can be linked at will, but it can only be handled properly, if the used vocabulary/ies are (1) known beforehand, or (2) intended interpretations can be inferred at run-time. In the later case, either current GML specifications required extension, or a purely RDF-based solution have to be developed. Final decisions and best practices are not (yet) available.

USING LINKED DATA TO COMBINE SDI WITH VGI

In our opinion, the linked data approach can be projected to SDI and it can be used as a possibility to combine SDI with VGI. In this work we advocate the support of linked data within spatial data infrastructure. In this section we successively suggest some solutions for linked data to classical SDI standards, and give examples of novel ways of data usage, detail the relation to VGI, and argue for benefits.

In general we foresee a projection of the notion of linked data (and related technology) to current SDI structures (right side of Figure 6). The basis of the proposed solution is that an SDI consists of several elements including people, datasets, technologies, standards and policies facilitating the sharing of data. Therefore, if the unit component of an SDI is considered to be a "resource" then relationships between resources become more visible. In this way, data provided by an SDI can be easily connected to other sources (left side of Figure 6). We will describe the overall approach, alternative strategies for applying the approach and its impacts on SDI in the following section. The proposed solution considers the application of RDF, which by definition is a standard for encoding definitions and relationships between resources. In this section we propose the following strategies for resource orientation:

- Full resource orientation through native RDF repositories.
- Semi-resource orientation through augmentation.
- Semi-resource orientation through mediator services.

All the strategies proposed for applying the proposed solution are resource-oriented; however, they differ in the degree of resource orientation. Thereafter we subsequently argue for possible adaptation of classical data provision paradigm (large rectangle in Figure 6) and for required changes in the use of services (small rectangle in Figure 6).

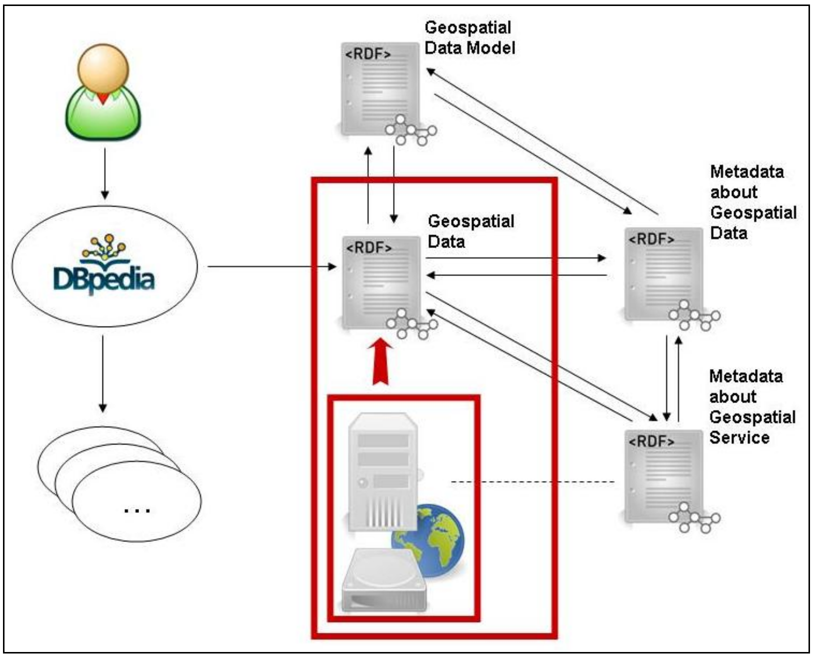

Figure 6. All strategies considered in one figure.

Full Resource-Orientation through Native RDF Repositories

The direct application of linked data to SDI attempts to drastically convert all (meta) data models and sets available in SDI into a single data structure. In parallel the naming schema in SDI is shifted from URNs to URIs. Services will be adopted accordingly. At the end, this would mean GML application schemas and metadata profiles will be encoded in RDF-S, and data and metadata sets will be directly represented in RDF. We present this approach in Figure 7, where examples are included in italics. The RDF will include additional links to other data, as for example links from DBpedia or to LinkedGeoData resources.

Assuming VGI is also represented in RDF, data from SDI and VGI can be linked and queried for relations. This solution may end the long-time discussion on distinguishing data from metadata as currently reflected in the use of GML and separate metadata documents, which for example use ISO 19115 and ISO 19139. Still links have to be established in the first place and maintained thereafter. Continuation of resources and links poses new challenges, because the joined data is not gathered by a small and well-known community any more. An unknown number of sources become connected in unpredictable ways. In a sense this supports the transition of the classical, cathedral-style of SDI to the bazaar-style that has been imposed by VGI (Budhathoki et al. 2008).

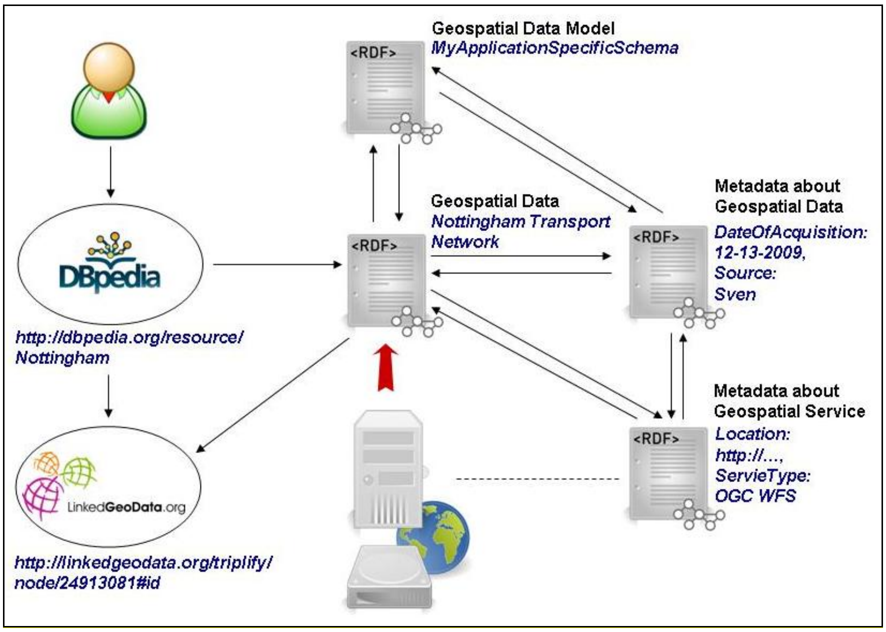

Figure 7. Turning SDI data and metadata into native RDF repositories

In our opinion this scenario is optimal in terms of linking available elements, but unrealistic in terms of pure data load. Storage is a central concern of this approach. It remains unclear where links between data from various sources should be stored and how distributed data can be queries in a semantic web fashion. Also, solutions for describing the intended interpretation of elements still have to be included. Real world comes with heterogeneity and this is also applicable to technology and information infrastructures. We believe that this strategy is inconceivable for many reasons. Current linked data publishers and providers represent data entities in the form of RDF triples. Links among such entities also maintain this pattern, where the subject of the triple is a source entity on a data source, the object of the triple is a target entity in other data source. The third piece, the predicate, determines the type of the links by using common vocabularies (like FOAF and SKOS (Miles and Bechhofer, 2009)), which define semantically the links among data entities. Although these typed relationships may span multiples data sources by just navigating across these explicit links, it seems to be an insufficient means to abstract (model) large scale issues like environmental sustainability that requires complex relationships among its components instead of just using the subject-object-predicate pattern. The real impediment, however, stems from the social, policies, collaborative aspects that are part of an SDI. While this approach would be technically plausible, in terms of policies, collaborations and arrangements would be an endless process.

Semi-Resource-Orientation through Augmentation

An alternative strategy is the augmentation of existing elements of SDI with references to external resources. Following this approach, the SDI community actually opens up to new horizons and provides native service interfaces that suit the linked data needs. This will be achievable if we consider an extended representation for geospatial data and services that includes and references to linked data. Logically we are talking about serving RDF, cross-referenced with established geospatial data formats such as GML and disseminated through current geospatial service interfaces. This implies mixing the current geospatial service architecture (defined by ISO19119) with a resource oriented architecture where appropriate. Candidates for resource oriented architectures include REpresentational State Transfer (REST, Fielding, 2000) and Web Services Resource Framework (WSRF, OASIS 2006). REST has gained popularity within the mobile and location based services community. WSRF standardizes the representation of stateful resources within service oriented architectures. Related studies have explored the integration of geospatial and grid computing through the import of geospatial data schema into WSRF service descriptions (Hobona et al, in press; Woolf and Shaon, 2009).

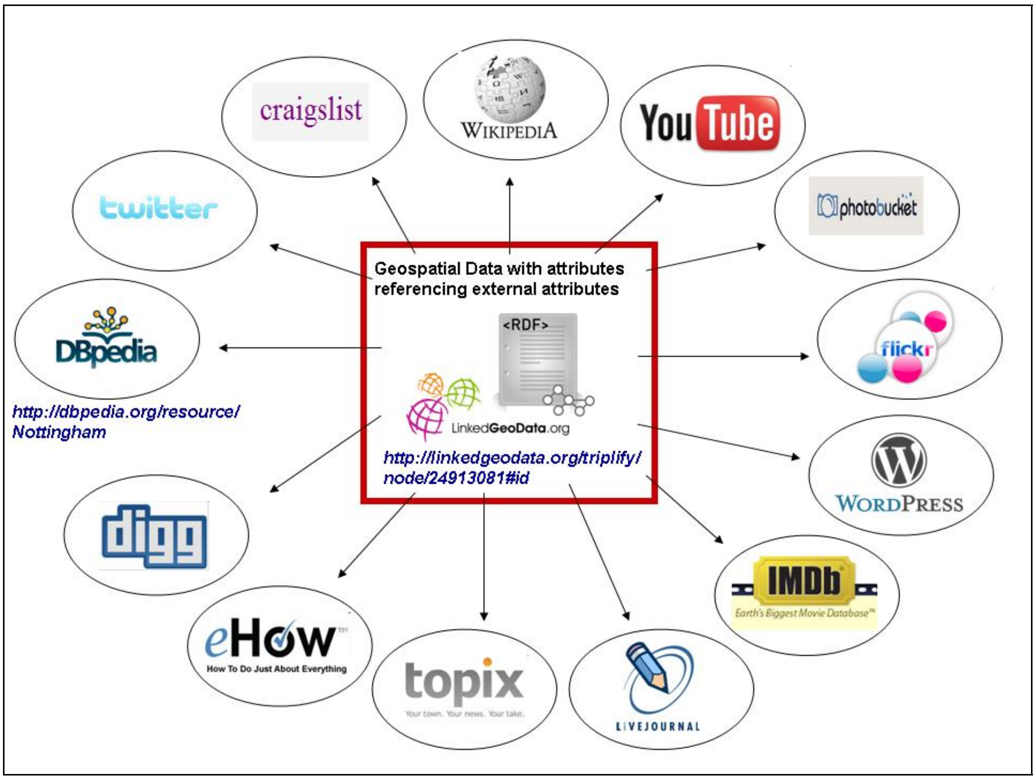

Figure 8. Augmentation of SDI elements

This approach leads to multiple benefits. First, linked data datasets are no longer static but dynamic, since structured content is generated on-the-fly, in a request basis. This allows user to access and discover up-to-date content. This approach also enables to convert VGI data into structured data where possible. There exist limitations as well. First, navigating or browsing capabilities present in linked data are greatly affected with dynamic structured content. Describing explicitly structured data in advance makes it possible to use sophisticated query capabilities as those provided by SPARQL query language.

Semi-Resource-Orientation through Mediator Services

Another alternative strategy is the use of mediators in applying resource orientation in SDI. Mediators offer content negotiation through the transformation of data in order to make it suitable for use in other applications (Wiederhold, 1992). Mediators have been applied in content negotiation for taxonomy-based information sources (Tzitzikas et al., 2005), therefore they could provide a 'bridge' between linked data and SDI if elements of SDI are uniquely identifiable. Current web services transparently make use of some capabilities for content negotiation in HTTP (Holtman and Mutz 1998) to allow client applications (like browsers) to negotiate various aspects of input and output data, for example media type, compression, character set and the language of a requested web page.

Assuming that a given resource has multiple representations associated with it, content negotiation (conneg) would allow a client to make preferences about which resource representation to retrieve from a transparently negotiable resource for the corresponding conneg-enabled service. Notably, within this approach classical SDIs can remain intact. Following the linked data approach, desired data encodings can be requested at runtime. Hence, clients may request geospatial data in (classical) GML and may retrieve (with little mortification of the request) the same data encoded in RDF, or any other format (Figure 9). This is a direct application of the notion of content negotiation, as already common in the web context (Holtman and Mutz 1998).

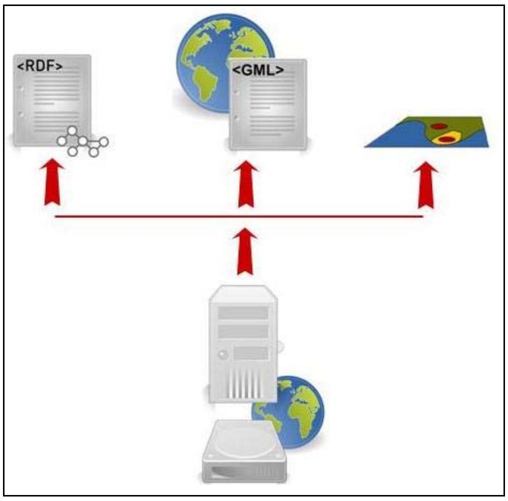

Figure 9. Mediation Service for content negotiation

A recent study by van de Sompel et al. (2009) proposed a novel application, called Memento, to add a time dimension on preservation and curation of data. The authors exploit the content negotiation characteristics to let users retrieve past versions of current web sites, envisioning somehow a time traveler for the web. This gives an idea of the potentials of content negotiation if were applied to conneg-enabled services within SDI.

The primary difference between this approach and the augmentation approach is that augmentation would require extension of current geospatial information models to include references to external resources. In contrast, mediators do not need the information models to be modified in advance. Instead, mediators would require an additional 'catalogue' for cross-referencing between geospatial data (or services) and external resources.

Discussion

Connecting conveniently data and metadata is a must to exploit the potential of current SDI catalogues and to link resources within SDI. In the SDI community, data and services are searchable by putting their metadata description in centralized catalogues. This strategy is appropriate for expert users because they perform elaborate queries using formalized metadata descriptors over a handful of well-documented repositories. Populating data catalogues in this way is a time-consuming task and it also requires experts in the concrete domain since the creation of metadata is still a manual, error-prone process, disconnected from the geospatial data assembly-line. Consequently, references to the physical location of datasets are often either missing or erroneously encoded. This limitation derives from the physical separation of data and metadata. Further, current geospatial data and metadata standards do not offer mechanisms for linking to external resources such as content from VGI sites. The previous section proposed some approaches for linking geospatial data and metadata to VGI through resource-oriented strategies. The envisioned result could be termed *Spatial Linked Data Infrastructure*.

Applying the notion of linked data to currently existing concepts on SDI metadata raises two concerns. First, should metadata contain links to other resources of the same SDI? Second, should metadata contain links to external resources (including elements of third-party SDIs)? Both concerns apply to metadata about a data set, but also to service metadata. Technically neither of these issues causes any problems, because recent metadata standards (ISO 2003; ISO 2005; ISO 2007; OGC 2005) allow for links using xlink to connect to external resources. The primary difference to the linked data approach is the encoding language. On the conceptual level, however, it becomes a question to the information community. Such community is usually defined as a group of people, which share a common interest and use a shared vocabulary, in which intended interpretations have been clarified within the community. In such cases, the SDI requires a component for resolving the identifiers of resources, to which pointers exist. Following the recent linked data suggestion of using URIs, such resolvers are provided by DNS.

Once clarified, how should we link from external resources to SDIs and especially to content managed within classical SDI catalogues? This concern does not directly affect the information community of the SDI, but the policy of making their content available to the outside. Again, resources within the SDI require identifiers and a common resolver has to be provided. Already there are indications of initiatives establishing registries for content other than data, for example, the INSPIRE Registry[xx] offers a feature concept dictionary and the Global Earth Observation System of Systems (GEOSS) offers a Standards Registry[xxi]. A common resolver of identifiers across these initiatives could help to address the linking of VGI as well. Any other activities are topic to the outer-SDI environment. Either pointers to clearly retrievable resources are used, or catalogue clients have to be established. A best practice, established by the international geospatial community, is currently missing.

So far, we considered the classical setting of SDI. With the linked data notion, we may even revolutionize this common setting by breaking the (anyway fuzzy) separation between data and metadata. Once we acknowledge the linked data principles, they can be directly applied to the current way of using the OGC and ISO standards for geospatial information. Each XML document, being GML, a metadata record following the ISO profile offered by a catalogue service based on the ebRIM profile, or any other element of SDI can make use of the xlink to connect to another resource inside the SDI. This way, data values can point to information about uncertainty, previous processing steps (also known as lineage) and any other piece of information in a unique manner. Of course, this suggests considering all elements of an SDI to be a resource. However, we contend that most, if not all, elements of SDI can be assigned unique identifiers such as URL (if they are accessible on the web) or URN (if they are not accessible on the web).

A much more general concern that applies to the overall linked data paradigm considers the data model used to encode the range of data links. Linking data does not specify the meaning of elements, which are pointed to. It makes only sense to link data in cases where we know the intended interpretation. The Nottingham use case provides already a rich set of examples. With GML, as well as RDF, links may point to anything. Structures used to represent this flexible content have to be communicated. This may either happen in a pre-deployment face, i.e. the meaning of used data structures is fixed and well known, or at run-time. Well known meaning can be defined by user requirements, or by detailed specifications using the context of the used links. Enabling run-time exploitation of data structures (or models) requires inclusion of data model information inside the link itself or to the range of the link. Approaches for developing a Spatial Linked Data Infrastructure may apply full resource orientation by using native RDF repositories as the only option for resource presentation. We already argue that this approach seems unrealistic. Instead we suggest semi-resource-oriented approaches, and favor mediation services that implement content negotiation. The final decisions have not been taken and reference implementations are required.

FUTURE RESEARCH DIRECTIONS

Having reviewed some open issued concerned with resource identification and representation and semantics of links, this section identifies future challenges and research directions that must be addressed to achieve the goal of connecting geospatial web services and SDI with linked data and VGI communities. Main areas of research include data population, discovery, access, and maintenance.

Publishing classical geo-data and services is currently covered by the CSW specification of OGC and so called Geoportals (Bernard et al., 2005). Moving to Spatial Linked Data Infrastructures requires revisiting at least the following core questions: How to publish resources? How to publish links? How to ensure links validness? How to propagate links updates?

As in the case of service interoperability between distributed infrastructures, as for instance SDI-Grid (Lee and Percivall, 2008), linked data poses new challenges for data services that the geospatial community should immediately address. Consuming linked datasets are, or should be, based on customized regrouping and restructuring the spectrum of data sources over the linked data cloud, according to specific user's perspective and needs. Relating the needs of the user to a dataset by considering the resources referenced by that dataset could offer a new and innovative approach for geographic data discovery. Further, the approaches in 'link analysis' could offer mechanisms for evaluating the relevance of linked geospatial datasets to the needs of a user (Borodin et al. 2005). One of the most successful applications of link analysis is the PageRank algorithm applied in Google (Brin and Page, 1998). Furthermore, in a previous section we proposed the establishment of a common resolver for identifiers. This presents a challenge for the SDI community to develop a service specification, conformant to ISO 19119, for resolving identifiers in support of both data and service discovery. Research questions for discovery in Spatial Linked Data Infrastructures would include: How to discover resources and services together? How to discover links? How to perform filtering capabilities? How to provide concrete views or perspectives of the whole liked data for specific user needs?

Future work will also have to consider the uniform access to heterogeneous sources from both SDI and VGI. This suggests a vision where geospatial data does not only offer the information it contains but also acts as a pointer towards other sources of information. Within service oriented architecture, a RESTful approach could help to implement this vision. However, it is necessary for the SDI community to consider how GIS and OGC web services will be able to access the heterogeneous data sources offered by VGI through a uniform and standardized mechanism. Again, we highlight that a solution would have to be developed through consensus, at an international level, in order to guarantee that the full potential of linked geospatial data is harnessed.

Establishing links between SDI and VGI resources could be a highly valuable capability. However, some of the safety critical applications of geospatial data will require the connections between linked data to be maintained through approaches such as versioning and caching. Versioning would allow changes in a referenced resource to be reported back to the dataset referencing the resource. The URN approach used by the OGC already includes an element for 'version numbers'. However, it is not yet clear whether a similar versioning approach could be extended to URLs of VGI content. Caching would allow for previous versions of resources to be stored and persisted, in case, they are required in the future. In summary, the challenges involved in the management of linked geospatial data will require additional research on how to propagate and report changes to links.

CONCLUSION

The purpose of metadata in SDI is to help users (like scientists, decision-makers, analysts, etc.) to fully understand resources in order to assess its usefulness or applicability in their daily tasks. In contrast, the aim of metadata in linked data (at the current stage) and above all in VGI content is to describe something that can be discovered and browsed easily. For instance, with regards to VGI content, people share videos, photos, and such types of resources, by tagging a resource with minimal descriptors: a title, key words or tags, a pointer to the very resource and, in some cases, a location reference. Simple metadata mechanisms like tagging are often sufficient for general discovery purposes, but clearly lack of suitable mechanisms for filtering information, setting query criteria and interpreting the search results, necessary for specific discovery tasks.

Geospatial data from linked data sources, VGI and SDI is addressed to different audience since they have different needs and requirements. This implies that metadata techniques are also quite different, since describing something that can be discovered easily is different from describing something that can be fully understood. However, this does not mean that these worlds need to keep separated since we can we use the existing links, tags and annotations to help users connect to other related geospatial data. In this sense, SDI content needs to exploit the vast amount of geospatial data present in social networks as for example, people who are interested Geonames place names around the city of Nottingham might also like to inspect street and elevation layers of this city from SDI catalogues services. Conversely, people who have accessed to some parcels layers for Nottingham might also like to know some Flickr pictures from locals in such areas. Similarly, knowing the network of users and organizations that have interest in common datasets could help in reinforcing social and professional ties among peer stakeholders.

Now, as the technology has been identified and a solution has been outlined, we moved closer to the third generation of SDI. We are aware of the existing practical and research challenges, and the foundations for reference implementations are set. It is up to us and the all other members of the community to implement a shared Spatial Linked Data Infrastructure.

REFERENCES

Ackland, R. (2009). Social Network Services as Data Sources and Platforms for e-Researching Social Networks. *Social Science Computer Review, 27*:481-492.

Alameh, N. (2003). Chaining Geographic Information Web Services. *IEEE Internet Computing, 7*(5), 22-29.

Auer, S., Lehmann, J., & Hellmann, S., (2009) LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In *Proceedings of the ISWC 2009*, LNCS 5823, pp. 731-746.

Becker, C., & Bizer, C. (2009). Exploring the Geospatial SemanticWeb with DBpedia Mobile. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 7*(4): 278-286.

Bernard, L., Kanellopoulos, L., Annoni, A., & Smits, P. (2005). The European Geoportal—one step towards the establishment of a European spatial data infrastructure. *Computers, Environment and Urban Systems, 29*(1), 15-31.

Bernard, L., & Craglia, M. (2005). SDI—from spatial data infrastructure to service driven infrastructure, research workshop on cross-learning between SDI and II. First research workshop on cross-learning on spatial data infrastructures (SDI) and information infrastructures (II), March31–April1. International Institute for Geo-Information Science and Earth Observation.

Berners-Lee, T., Fielding, R., & Masinter, L. (1998). Uniform Resource Identifiers (URI): Generic Syntax. Internet Engineering Task Force (IETF) Memo – RFC 2396.

Bishr, M., & Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth *GeoJournal*, *72*(3-4), 229-237.

Bizer, C. (2009).The Emerging Web of Linked Data. *IEEE Intelligent Systems*, *24*(5): 87-92

Bizer, C., Heath, T., & Berners-Lee, T. (2009a). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, *5*(3): 1-22.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009b). DBpedia - A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, *7*(3): 154-165.

Borodin, A., Roberts, G. O., Rosenthal, J. S., & Tsaparas, P. (2005). Link analysis ranking: algorithms, theory, and experiments. ACM Transactions on Internet Technology (TOIT), 5(1), 231 - 297.

Brickley, D., & Guha, R.V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004

Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine, 7th International World Wide Web Conference (Vol. 30, pp. 107-117).

Budhathoki, N.R., Bruce, B., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal: An International Journal on Geography*, *72*:149-160.

Callahan, K. M. (1985). Social Science Research in the Information Age: Online Databases for Social Scientists. *Social Science Computer Review*, *3*: 28-44.

Crompvoets, J., and Bregt, A. (2007).National spatial data clearinghouses, 2000–2005. In H.Onsrud (Ed.), Research and theory in advancing spatial data infrastructure (pp.133–146).Redlands: ESRI Press.

Díaz, L., Granell, C., & Gould, M. (2009). Spatial Data Integration over the Web. In V. E. Ferraggine, J.H. Doorn, L.C. Rivero (Eds.), *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 325-333). Hershey, PA: Information Science Publishing.

European Parliament and Council (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal on the European Parliament and of the Council.

Executive Office of the President (1994). Coordinating geographic data acquisition and access: The national spatial data infrastructure. Executive order 12906, Federal Register 59.

Fielding, R. (2000). Architectural Styles and the Design of Network-based Software Architectures. PhD Dissertation, University of California, Irvine, http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

Gahegan, M., Luo, J., Weaver, S.D., Pike, W., & Banchuen, T. (2009). Connecting GEON: Making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure. *Computers & Geosciences*, *35*(4): 836-854.

Goodchild, M.F., Egenhofer, M., Fegeas, R., & Kottman, C. (1999). *Interoperating Geographic Information Systems*. Norwell: Kluwer Academic Publishers.

Goodchild, M.F. (2007). Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, *2*: 24-32.

Goodchild, M.F. (2009). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, *3*(2): 82-96.

Guarino, N. (1998). Formal Ontology and Information Systems. in: N. Guarino (Ed.), Formal Ontology in Information Systems (FOIS'98).

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing 7*(4): 12-18.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D 42*: 335-346.

Hill, L.L. (2006). *Georeferencing: The Geographic Associations of Information*. Digital Libraries and Electronic Publishing. The MIT Press.

Hobona, G., Fairbairn, D., Hide, H., & James, P. (in press). Orchestration of Grid-Enabled Geospatial Web Services in Geoscientific Workflows. *IEEE Transactions on Automation Science and Engineering.*

Holtman, K., & Mutz, A. (1998). Transparent Content Negotiation in HTTP. Internet Engineering Task Force (IETF) Memo – RFC 2295.

Hubmann-Haidvogel, A., Scharl, A., & Weichselbraun, A. (2009). Multiple coordinated views for searching and navigating Web content repositories. *Information Sciences 179*: 1813-1021.

INSPIRE (2009). INSPIRE Data Specification Transport Networks (Version 3.0). INSPIRE Thematic Working Group on Transport Networks.

ISO (2003). 19115 Geographic Information - Metadata. ISO/TC211 Standards.

ISO (2005). 19119 Geographic Information - Services. ISO/TC211 Standards.

ISO (2007). 19139 Geographic information - Metadata - XML schema implementation. ISO/TC211 Standards.

Janowicz, K., & Keßler, C. (2008): The Role of Ontology in Improving Gazetteer Interaction. *International Journal of Geographical Information Science, 22*(10): 1129-1157.

Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maue, P., & Stasch, C. (in press). Semantic Enablement for Spatial Data Infrastructures. *Transactions in GIS.*

Klyne, G., & Carroll, J.J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 February 2004

Lassila, O., & Swick, R. (1999). Resource Description Framework (RDF) Model and Syntax Specification. Retrieved (20.10.2009) from http://www.w3.org/TR/PR-rdf-syntax/.

Lee, C., & Percivall, G. (2008). Standards-Based Computing Capabilities for Distributed Geospatial Applications. *Computer*, *41*(11): 50–57.

Lucchi, R., Millot, M., & Elfers, C. (2008). Resource oriented architecture and REST: Assessment of impact and advantages on INSPIRE. JCR Scientific and Technical Report. EUR 23397.

Masser, I. (1999). All shapes and sizes: The first generation of national spatial data infrastructures. *International Journal of Geographical Information Science*, *13*(1), 67-84.

Masser, I. (2007). Building European spatial data infrastructures. Redlands, CA: ESRI Press.

Maué, P., Schade, S., & Duchesne, P. (2009). OGC Discussion Paper 08-167r1: Semantic annotations in OGC standards. Technical report, OGC, 2009.

Miles, A., & Bechhofer, S. (2009). SKOS Simple Knowledge Organization System Namespace Document - HTML Variant. 18 August 2009 Recommendation Edition.

Nebert, D. (2004). SDI Cookbook, Version 2.0.

OASIS (2006). Web Service Resource 1.2 (WS-Resource). OASIS Standard. http://docs.oasis-open.org/wsrf/wsrf-ws_resource-1.2-spec-os.pdf

OGC (2005). OpenGIS® Filter Encoding Implementation Specification, Version 1.1. The Open Geospatial Consortium.

OGC (2006). OGC White Paper - An Introduction to GeoRSS: A Standards Based Approach for Geo-enabling RSS feeds. The Open Geospatial Consortium.

OGC (2007a). OpenGIS® Geography Markup Language (GML) Encoding Standard - Version 3.2.1. The Open Geospatial Consortium.

OGC (2007b). OpenGIS® Catalogue Service Implementation Specification (Version 2.0.2). Open Geospatial Consortium.

OGC (2007c). OGC® KML (Version 2.2). Open Geospatial Consortium.

Omran, E.L.E., & van Etten, J. (2007). Spatial-data sharing: Applying social-network analysis to study individual and collective behaviour. *International Journal of Geographical Information Science*, *21*(6), 699-714.

Onsrud, H., Poore, B., Rugg, R.T., & Wiggins, L. (2004). The future of the spatial information infrastructure. In R.B. McMaster and E.L. Usery (Eds.), A Research Agenda for Geographic Information Science (pp.225–255). Boca Raton: CRC Press.

van Oort, P. A. J., Hazeu, G. W. , Kramer, H., Bregt, A.K., & Rip, F. I. (in press). Social networks in spatial data infrastructures. *GeoJournal*.

Rajabifard, A., Binns, A., Masser, I., & Williamson, I. (2006). The role of sub-national government and the private sector in future spatial data infrastructures. *International Journal of Geographical Information Science*, *20*(7): 727-741.

Rajabifard, A., Kalantari, M., & Binns, A. (2009). SDI and Metadata Entry and Updating Tools. In B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Eds.), *SDI Convergence. Research, Emerging Tools, and Critical Assessment* (pp. 121-135). Rotterdam: Netherlands Geodetic Commission.

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

Riehle, D. (2007). The Economic Motivation of Open Source Software: Stakeholder Perspectives. *IEEE Computer*, *40*(4), 25-32.

van de Sompel, H., Nelson, M.L., Sanderson, R., Balakireva, L.L., Ainsworth, S., & and Shankar, H. (2009). Memento: Time Travel for the Web. In arXiv.org, 6 Nov 2009, http://arxiv.org/abs/0911.1112v2

Suda, B. (2006). Using Microformats. O'Reilly Media.

Tzitzikas, Y., Spyratos, N., & Constantopoulos, P. (2005). Mediators over taxonomy-based information sources. *The International Journal on Very Large Data Bases*, *14*(1), 112-136

Wiederhold, G. (1992). Mediators in the Architecture of Future Information Systems. *IEEE Computer Magazine*, *25*(3): 38-49.

Woolf, A., & Shaon, A. (2009). An Approach To Encapsulation Of Grid Processing Within An OGC Web Processing Service. GIS.Science, 3, 82-88.

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

ADDITIONAL READING SECTION

Abdelmoty, A. I., Smart, P. D., Jones, C. B., Fu, G., & Finch, D. (2005). A critical evaluation of ontology languages for geographic information retrieval on the Internet. Journal of Visual Languages & Computing, 16(4), 269-382.

Andrei, M., A.J. Berre et al. (2008) Swing: An integrated environment for geospatial semantic web services. In S. Bechhofer, M. Hauswirth, et al. (Eds.), 5th European Semantic Web Conference (ESWC2008), LNCS5021, pp 767-771. Springer, 2008.

Berners-Lee, T., J. Hendler and O. Lassila (2001). The Semantic Web. *Scientific American Magazine*, *284*(5): 34-43.

Egenhoger, M.J. (2002). Toward the semantic geospatial web. Proceedings of the 10th ACM international symposium on Advances in geographic information systems, 1-4, Virginia, USA.

Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4): 211-221.

Goodchild, M.F., Fu, P., & Rich, P. (2007). Sharing geographic information: An assessment of the geospatial one-stop. *Annals of the Association of American Geographers, 97*(2): 250-266.

Lemmens, R., Wytzisk, A., de By, R., Granell, C., Gould, M., & van Oosterom, P. (2006). Integrating Semantic and Syntactic Descriptions to Chain Geographic Services. *IEEE Internet Computing*, *10*(5), 42-52

Lieberman, J., T. Pehle et al. (2006). Geospatial Semantic Web Interoperability Experiment Report. Technical report, Open Geospatial Consortium.

de Longueville, B., N. Ostlander et al. (2009). Addressing Vagueness in Volunteered Geographic Information (VGI) - A Case Study. 11th International GSDI Conference, Rotterdam, Netherlands.

Probst, F., & Lutz, M. (2004). Giving Meaning to GI Web Service Descriptions. In 2nd International Workshop on Web Services: Modeling, Architecture and Infrastructure (WSMAI2004), Porto, Portugal, 2004.

Vickery, G. and S. Wunsch-Vincent (2007). Participative Web And User-Created Content: Web 2.0 Wikis and Social Networking. Organization for Economic.

W3C (2001). URIs, URLs, and URNs: Clarifications and Recommendations 1.0. Report from the joint W3C/IETF URI Planning Interest Group. W3C Note 21 September 2001. http://www.w3.org/TR/uri-clarification/.

Weets, G. (2006). Toward a Single European Information Space for Environment. ESA - DG INFSO Interoperability workshop: Architecture workshop in support of GEO and GMES. Frascati, Italy.

Yue, P., Gong, J., Di, L., He, L., & Wei, Y. (in press). Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure. *GeoInformatica*.

KEY TERMS & DEFINITIONS

Discovery: The action or task of searching or finding for geospatial datasets, metadata, services and resources in general.
Linked data: Represents a set of structured data sources over the Internet connected via typed relationships.
Metadata: A set of information descriptors used to describe the characteristics (format, quality, use, proprietary, etc.) of geospatial datasets, services and resources in general.
Retrieval: The action or task of gathering or accessing to geospatial data sets, services and resources in general.
Spatial data: Data sets that refer explicitly to a geospatial connotation and/or are georeferenced.
Spatial Data Infrastructure (SDI): Is a type of information infrastructure for enhancing geospatial data sharing and access.
Spatial Linked Data Infrastructure: The idea of projecting linked data principles to SDI architecture and standards so that SDI and VGI content can be easily combined.
Spatial Web Service: Is a software component that delivers and processes any geospatial data over the Internet.
Volunteered Geographic Information (VGI): Is the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007).

[i] http://www.w3.org/DesignIssues/LinkedData.html
[ii] http://www.csc.fi/english/pages/parade
[iii] http://www.geonames.org
[iv] http://linkedgeodata.org
[v] http://www.rdfabout.com/demo/census/
[vi] http://www.dbis.informatik.uni-goettingen.de/Mondial/
[vii] http://riese.joanneum.at/
[viii] http://europa.eu/estatref/download/everybody/
[ix] http://data.ordnancesurvey.co.uk
[x] http://telegraphis.net/data/
[xi] http://ontologi.es/rail/
[xii] http://www.eionet.europa.eu/gemet/rdf?langcode=en
[xiii] http://jsexplicit.sourceforge.net
[xiv] http://iws.seu.edu.cn/services/falcons/objectsearch/index.jsp (accessed 10th November 2009).
[xv] http://dbpedia.org (accessed 10th November, 2009).
[xvi] http://ode.openlinksw.com/ (accessed 10th November, 2009)
[xvii] http://www.geonames.org/ontology/ (accessed 11th November, 2009)
[xviii] http://xmlns.com/foaf/spec
[xix] http://portal.opengeospatial.org/files/?artifact_id=24045
[xx] http://inspire-registry.jrc.ec.europa.eu/
[xxi] http://www.earthobservations.org/gci_sr.shtml