Spatial Data Infrastructures and the Semantic Web of Spatial Things in Australia

Research opportunities in SDI and the Semantic Web

David A. McMeekin
Cooperative Research Centre for Spatial Information
Department of Spatial Sciences
Curtin University
Perth, Australia
d.mcmeekin@curtin.edu.au

Geoff West
Cooperative Research Centre for Spatial Information
Department of Spatial Sciences
Curtin University
Perth, Australia
g.west@curtin.edu.au

Abstract—**Spatial Data Infrastructures have recently become a crucial part of national infrastructures. Example users are governments using them to make informed policy decisions and the private sector using them in order to understand their customers better. It is estimated that Australian spatial industry revenue is in excess of $1.35 billion annually. The Cooperative Research Centre for Spatial Information in Australia is currently working with the Commonwealth, jurisdictions and the private sector to understand the research required in this area to support both the public and private sectors in their decision and policy making based upon these infrastructures. This paper presents an early perspective as to possible research areas in this field. An underlying theme that reoccurs in the research is the need to consider usability of such systems and the need to move beyond just data to orchestration of processes to obtain derived products.**

Keywords-component; sdi; semantic web; ontology; spatial data; rdf.

I. INTRODUCTION

An increasingly important component of national infrastructure is Spatial Data Infrastructure (SDI) that consists of spatial information such as road centre lines, land parcels and land height data. SDIs are providing resources for policy decisions, emergency planning and governing as well as for private industry business decisions [1, 2]. In the years 2006 – 2007 it was estimated that direct revenue from the spatial industry in Australia was up to $1.35 billion and its use within other areas in Australia contributed $6.43 - $12.57 billion to the economy [22]. These factors have played a significant role in seeing the number of SDIs increasing rapidly worldwide [3, 4, 5, 6, 7].

The Australian Federal Government established the Office of Spatial Policy (OSP) in July 2011 that is "…a central policy unit, responsible for facilitating and coordinating spatial data management across Australian Government agencies… in response to outcomes of various reviews…" [8]. A challenge facing SDIs is not an information shortage but quite the opposite, the abundance of available information. In the modern world, apart from traditional quite static spatial data (cadaster, roads), much data is being collected continuously. There are geo-sensors, automotive traffic sensors, maritime traffic sensors, aerospace traffic sensors, weather sensors, seismic sensors to name just a few. There is also the constant supply of information (some of which is geo-tagged) through social network feeds such as Twitter and Facebook.

The data is abundant, but it is the ability to search and extract meaningful and useful information in appropriate time frames where many research challenges occur. Using the landscape of Australian SDIs as a backdrop this paper examines SDIs and the possibility of how the Semantic Web may be used to solve issues with making the vast array of data available in useful, meaningful ways to government, private industry and the general public.

II. WHAT IS A SDI

Commonly cited SDI definitions are: "a coordinated series of agreements on technology standards, institutional arrangements, and policies that enable the discovery and use of geospatial information by users and for purposes other than those it was created for" [9] or "…a framework of technologies, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve the use of geospatial data across multiple public and private organizations" [10]. Put another way, an SDI is the infrastructure required to make geospatial data available for use by those who *may or may not* be experts in the geospatial domain.

The Australian Cooperative Research Centre for Spatial Information (crcsi) is currently examining the SDIs available in Australia and New Zealand in order to facilitate coordinated research in the field within the two countries. Fig. 1 depicts a, not yet completed, picture of this landscape showing those organisations that have some involvement in SDI generation and control. The figure shows that there are many different groups involved in SDI development and usage and exploration reveals a certain amount of disconnectedness and duplication amongst the various organisations because of different funding sources (research *vs* jurisdictional SDIs for example).

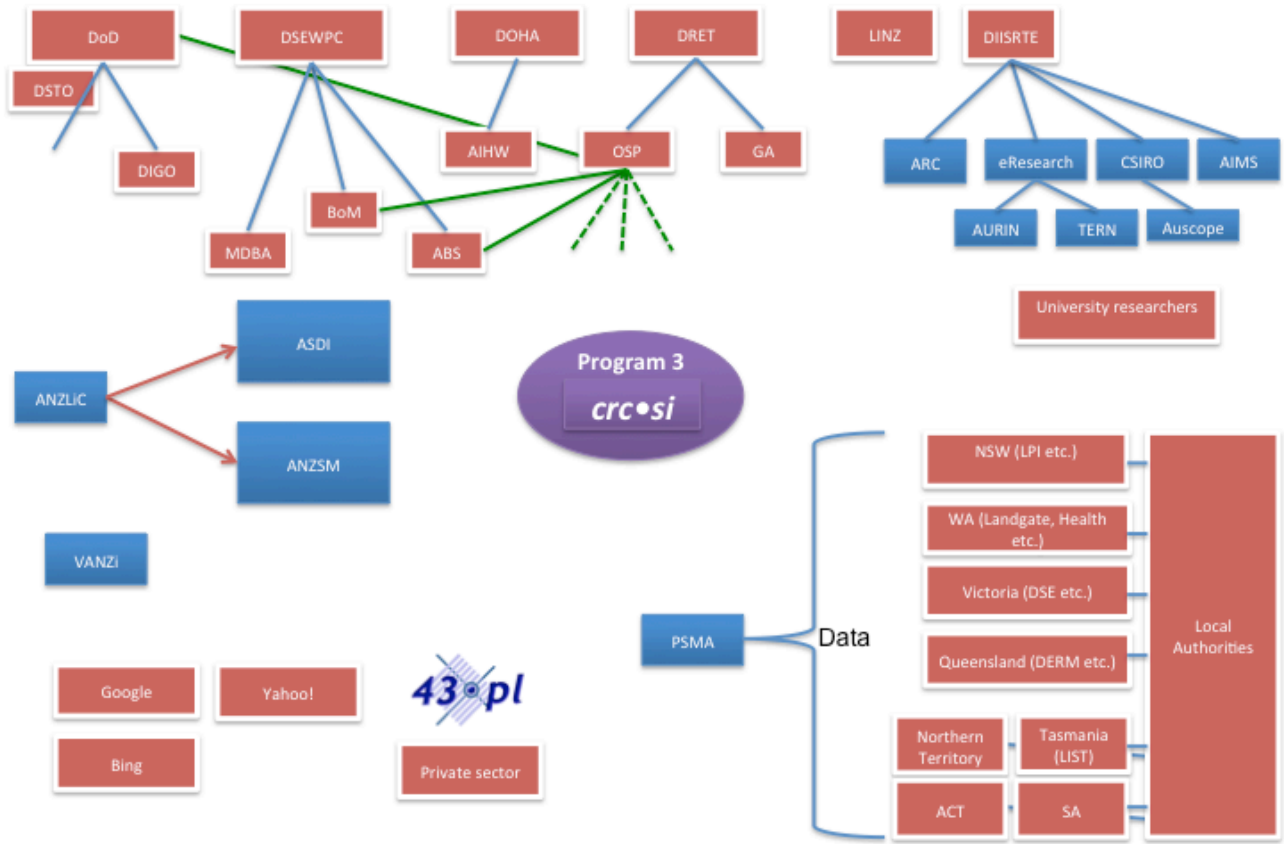

Figure 1. A partial diagram of the SDI landscape of Australia and New Zealand.

Presentations and interviews with more than 35 leading Australian and New Zealand experts from more than 25 different companies and government departments, has revealed the general consensus that within Australia and New Zealand, SDI development is occurring in isolated silos. In some cases, different agencies and organisations were completely unaware of the existence of other groups carrying out similar work.

As different organisations work in isolation of each other, often the data that is collected is stored in proprietary formats that are closely coupled with the specific software and hardware used to collect, store and process. This means that the collected data is of limited use to another group unless they have that same software. Therefore the ability and opportunity for data re-use is either low or even non-existent. This can be especially problematic and expensive for governments as information and understanding that may be gleaned from that data to better inform policy decisions is not possible and for them to go out and collect the required data can be very expensive.

The obvious solution to this scenario is for the data to be shared and available to others. Depending on the data collector this provision could be for a fee or provided free of charge. However, as already mentioned this is problematic when the data is stored in proprietary formats requiring specific software solutions. This is being addressed somewhat by the OGC (www.opengeospatial.org) with a number of standards for data and processes. However the effective discovery and interoperability of heterogeneous data sets is still an underlying problem in SDIs. The use of the Semantic Web Technology Stack, Fig. 2, is being proposed as one possible means to assist with solving this problem in a semi-automated or automated way.

III. SEMANTIC WEB

Berners-Lee et al. [11] published, what is now the seminal paper for the Semantic Web vision, where Peter and Lucy organize their mother's specialist appointments using their semantic web agents. The promotion of the Semantic Web by this article along with several others has increased Web users' expectations of what should be achievable on the Web [14]. The technologies described in that seminal article have matured over the past 11 years and the Semantic Web is now evolving steadily into a reality.

Feigenbaum et al. [12] present two case studies in which the Semantic Web is effectively used: the first one in drug

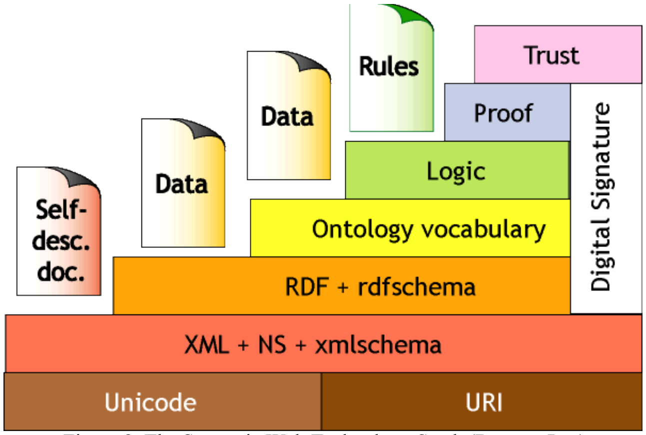

Figure 2. The Semantic Web Technology Stack (Berners-Lee).

discovery and the second one in health care. These are two research areas in which it is vital to be able to meaningfully search large amounts of data. A simple search does not usually return effectively meaningful data. However, where these data sets have been created using Semantic Web technologies, searching these data sets is now returning effective useful results.

Fig. 2 shows the Semantic Web Technology Stack as proposed in a presentation by Berners-Lee to the W3C consortium. The Semantic Web is not a different web to the current one, but simply the natural evolution of the traditional web. On the traditional web, the main focus is on making data available for humans to read. Although easy for humans to read, this kind of data is not easy for computers to automatically process. Hence, on the Semantic Web data is created in a way that makes it easy for computers to automatically process [11] and the technology stack shown in Fig. 2 enables computers to automatically process the data. From the human reader's perspective, nothing looks different. This is a major strength of the Semantic Web Technology Stack.

A concern with the Semantic Web is that for the data to be machine readable, it must all follow the same data structure. This is difficult because within the spatial data domain many large data sets already exist and to reformat them would be extremely expensive. In Australia there is also no government policy regarding the format that the data needs to be stored in for ease of access. This is quite different from the European situation where the European Union established a standard data structure that all European nations must comply with by 2019 [24]. This is to allow easy access to the data for different parties.

EuroGEOSS has recognized that the diversity of scientific and technology practices across the research community makes it close to impossible to impose a single solution for this interoperability. Hence, they have created a system that uses repositories and a brokering type system [15].

However, the Semantic Web is not attempting to have all data converted to a specific format, rather it is looking at having underlying standards available that data on the web can embrace, therefore enabling machines to be able to process the data to extract meaningful and insightful knowledge from it without compromising the data's eloquence and uniqueness [14].

Fig. 3a displays one view of the evolution of the web and the approximate place on the time scale where we believe SDIs are currently positioned. Fig. 3b shows where we hope SDIs will be in the next five years. In consultation with the different experts throughout Australia and New Zealand there was a minority who thought the current location of SDIs on the time scale was a little too far up. For instance, ftp is still a major technology used to move data around. Also, where SDIs will be in five years was looked at as being slightly over ambitious by some. Meeting with that many different experts and academics who in the main had only small differences of opinions as to where SDIs are and will be indicated that their location on the diagram may reflect reasonably accurate positioning and aspiration.

IV. SEMANTIC WEB TECHNOLGIES

An important technology in the Semantic Web for assisting with data is ontology. Ontology stems from ancient philosophy with Parmenides being one of the earliest philosophers to be credited with using ontology to characterize the nature of reality. Software systems [16] was one of the earliest adopters and promoters of ontology use. It stemmed from within the Artificial Intelligence (AI) community to formally represent knowledge within AI systems to allow reasoning and inference. Ontology was defined as "… an explicit specification of a conceptualization" [16]. The W3C Web Ontology Language Overview defines it as "the representation of meaning of terms in vocabularies and the relationships between those terms" [17]. It is through these relationships and representations that different data that describes similar entities can be linked to form new knowledge.

Other technology within the Semantic Web are Resource Description Frameworks (RDF) and Resource Description Framework Schemas (RDF/S). RDF is an XML based language that is used to describe different resources. It is used as a standalone metadata document that describes the document itself [18]. An RDF can be searched by an application or web service so that it can build an understanding about what the document contains. From the understanding that it builds, depending on what its purpose is, it may decide to use or not use that document.

Ontologies and RDFs are just two technologies that are being used within the Semantic Web. There are several other technologies such as Vocabularies and Folksonomies but this is by no means an exhaustive list.

V. Spatial Marketplace

The Australian New Zealand Land Information Council (ANZLIC) has commissioned an exciting new venture in the development of the Spatial Marketplace. This development recognizes that government data needs to be unlocked and private spatial data suppliers can place their data and processes for others to use either for free or fee. It is to be a place that spatial data consumers can come to in order to find and use the provided data and processes [21, 22]. One operating principle of the marketplace is that "Everyone can discover and anyone can play" [19].

A major challenge that the spatial market place faces is how do consumers get the information that they need from it. A search feature is the most obvious way but a standard search will not bring in related data. Spatial data sets can be extremely large, sometimes terabytes in size, hence search is important but the need to have smart search facilities is even more important. When searching for information to assist with making decisions, often valuable information is not that which is returned from a search but the combination of different information. To get the combination usually multiple searches need to occur.

A Semantic Web search approach [23] through the Spatial Marketplace may be a solution that in a single search the data requested as well as extra related data is returned thus providing the searcher with enough information to make an informed decision or in the case of a policy maker, provide enough information for them to continue gathering the needed information for policy decision making.

Ontologies and RDFs allow the creation of Smart Data. Smart Data is data that contains enough information for computers to be able to make inferences from the data. The ontologies and RDFs along with other technologies such as rules allow new data and knowledge to be inferred from the existing data being searched. Deploying this search capability into the spatial marketplace will create an effective spatial data repository for public and private use.

For a spatial data supplier a spatial marketplace provides a location for distributing their data either for fee or free. It also provides opportunity for developers to build applications and processes that would take advantage of the available data providing value added services to the community.

VI. Research Areas

A Semantic Web based spatial marketplace is a real possibility. However for this to happen there is still several areas that need to be researched. From consulting the different experts and organizations throughout Australia and New Zealand there were several research areas that repeatedly arose that need to be addressed.

The three over arching areas were a) Web-based architectures, b) Automation and c) Usability.

A. Web-based Architectures

For web-based architectures there are a growing number of standards developing, especially through the Open Geospatial Consortium (OGC). These standard include Web Processing Service (WPS), Web Feature Service (WFS), Keyhole Markup Language (KML), to name just a few. These standards are always open for modification and new standards are being invented as required. In Australia and New Zealand, ANZLIC recognises that the standards need to be explored in the Australian domain and is in the process of starting their own OGC forum to ensure that the needs of Australia and New Zealand are taken into consideration as these standards are developed and evolve.

B. Automation

Currently most of the creation of ontologies, RDFs and vocabularies is carried out manually.

Figure 3a. The current state of SDIs and the technologies used in them.

Figure 3b. The predicted state of SDIs in 5 years and the technologies they use.

This usually requires an expert in the field to go through the data and understand its semantics and meaning to manually create these artifacts. Automatically generating these types of artifacts through AI and machine learning is an area that has had some but not enough research activity.

Another research area is in the Federated Data Models domain. Currently much of the research has not closely examined federated data models but rather it has looked into non-federated data models. Many organizations have spent large sums of money in collecting spatial data and are unwilling to process the model to change the schema or data model because this can require much manual interpretation to map concepts and data types. Rather, a better approach is the creation of a community based schema which can be placed in front of the data sets that already exist. That which faces out from the data set to the world would be a community developed schema that enables web-based processes to simply access the data without modifying the original data

Supply chain generation, and reconciling licensing and terms of use are also areas of automation that requires research. In a spatial marketplace customers may search for information that ends up being supplied through various different suppliers but is delivered by a single web service provider. In this situation different data sets may have different licensing conditions and hence the final derived produce will be restricted within the different licenses.

C. Usability

System usability is an extremely important factor for the success of SDIs and the Spatial Marketplace. Interfaces by

Apple® and Google® are simple to use and hide the complexity of processing from the user. Usability does seem to be quite often overlooked in SDIs. In developing Semantic Web based applications, especially those commissioned by non GIS and spatial data experts [20] it is clearly pointed out that during the specification creation as well as at deployment time, client expectations need to be carefully managed. During consultations with the various experts and organisations throughout Australia and New Zealand this sentiment was echoed. Managing expectations as well as providing services that are usable for both expert and non-expert is important.

VII. Conclusion

Semantic Web provides the next step in the evolution of the current web. The realization of that which Berners-Lee described in 2001 is emerging. The area of spatial data is also an area evolving to be one of the most important data areas due to almost all data having some kind of location associated with it. For the Semantic Web for spatial information to live up to its full potential there is much research that needs to be carried out in order to provide the theoretical and practical underpinnings for the technologies that will be used in the coming years.

Acknowledgment

The work has been supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealth's Cooperative Research Centres Programme.

References

[1] Feeney, M.-E.. 2003. SDIs and decision support. In Developing Spatial Data Infrastructures: From concept to reality, eds. Ian Williamson, Abbas Rajabifard, and Mary-Ellen F. Feeney, 195-210. Boca Raton: CRC Press.

[2] Groot, R. 2001. Reform of government and the future performance of national surveys. Computers, Environment and Urban Systems 25 (4-5): 367-87.

[3] Crompvoets, J., A. Bregt, A. Rajabifard, & I. Williamson. 2004. Assessing the worldwide developments of national spatial data clearinghouses. International Journal of Geographical Information Science 18 (7): 665-89.

[4] Masser, I. 2005a. GIS Worlds: Creating Spatial Data Infrastructures, 1st ed. Redlands, CA: ESRI Press.

[5] Onsrud, H. J. 1998. Compiled Responses by Questions for Selected Questions: Survey of National Spatial Data Infrastructures. Retrieved May 18, 2012, from http://www.spatial.maine.edu/~onsrud/gsdi/Selected.html.

[6] McCall, M. K. 2003. Seeking good governance in participatory-GIS: A review of processes and governance dimensions in applying GIS to participatory spatial planning. Habitat International 27 (4): 549-73.

[7] Mennecke, B. E. 1997. Understanding the role of geographic information technologies in business: Applications and research directions. Journal of Geographic Information and Decision Analysis 1 (1): 44-68.

[8] Office of Spatial Policy. Retrieved May 18, 2012, http://www.ret.gov.au/Department/osp/Pages/OfficeSpatialPolicy.aspx.

[9] Kuhn, W. 2005. Introduction to Spatial Data Infrastructures. Presentation held on March 14 2005.

[10] What Is an SDI? - ArcNews Summer 2010 Issue. Retrieved May 18, 2012, http://www.esri.com/news/arcnews/summer10articles/what-is-sdi.html.

[11] Berners-Lee, T., Hendler, J., Lassila, O., The Semantic Web, Scientific American, 2001.

[12] Feigenbaum, L. Herman, I. Hongsermeier, T. Neumann, E. & Susie Stephens, The Semantic Web in Action, 2007.

[13] Kornack, D. & Rakic, P. "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130.

[14] Sheth, A., Ramakrishnan, C. & Thomas, C., 2005, Semantics for the Semantic Web: The Implicit, the Formal and the Powerful, International Journal on Semantic Web and Information Systems (IJSWIS), 1(1), pp. 1-18.

[15] New Tool to Help Monitor the Environment | Earthzine. Retrieved May 10, 2012, http://www.earthzine.org/2012/03/19/new-tool-to-help-monitor-the-environment.

[16] Gruber, T. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," Int'l J. Human-Computer Studies, vol. 43, nos. 5-6, 1995, pp. 907-928.

[17] OWL Web Ontology Language Overview. Retrieved May 18, 2012, http://www.w3.org/TR/owl-features/.

[18] Daconta, M. C, Obrst, L. J. & Smith, K. T. The Semantic Web : A Guide to the Future of XML, Web Services, and Knowledge Management. illustrated ed. Indianapolis, Ind.: Wiley Pub., 2003.

[19] ANZLIC: Australia New Zealand Spatial Market Place - ANZSM. Retrieved May 18, 2012, from http://www.anzlic.org.au/Infrastructures/Australia+New+Zealand+Spatial+Market+Place+-+ANZSM/default.aspx.

[20] Commandeur, T. J. , Asperen, P.C.M. van, 2008. The effect of Google Earth on the development of a SDI at a governmental agency, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4. Beijing.

[21] Johnson, J., & Crooks, C., 2012, Geonode and the Australian and New Zealand Spatial Marketplace: An Evolutionary Step in the Discovery, Access and Utilization of Spatial Resources, Global Spatial Data Infrastrucutre 13(GSDI 13), Quebec, Canada.

[22] The Value of Spatial Information - the Impact of Modern Spatial Information Technologies on the Australian Economy. (2008). (Original work published 2008) Retrieved from http://crcsi.com.au/Documents/ACILTasmanReport_full.aspx)

[23] Wilkinson, M., Vandervalk, B., & McCarthy, L., 2011, The Semantic Automated Discovery and Integradion (SADI) Web service Design-Pattern, API and Reference Implementation, Journal of Biomedical Semantics, 2 (8).

[24] 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).