

SPATIAL DATABASE CHALLENGES FOR FRAMEWORK DATA AND LOCATION-BASED SERVICES

Mike Jackson^a, and Peter Woodsford^{b*}

^aCentre for Geospatial Science, University of Nottingham, UK – mike.jackson@nottingham.ac.uk

^bSpatial, Cambridge, UK - consultancy.woodsford@ntlworld.com

Theme Sessions, TS-25: ThS-9

KEY WORDS: Spatial Database Management, Service Oriented Architecture, Framework Data, Location-Based Services, Interoperability, Update, Quality, Test-beds

ABSTRACT:

Mass-market location-based services (LBS) have had a number of false dawns. However, the belief in their potential revenue earning capacity to business and their prospective benefits to social applications remains. As a result, research and investment has continued into the resolution of the factors that have inhibited the success of services. The demands on data provision, spatial database management systems, dissemination and maintenance services in mass-market LBS have not yet been put under operational stress as other technological limitations have proved to be stumbling-blocks at earlier stages in the development chain. Rapid progress in the resolution of these inhibitors is, however, being made. It is therefore important that data and data management issues do not become a further brake on a successful roll-out of mobile LBS. At the same time important advances are occurring in the management and update of framework data. The paper examines some of the key developments and drivers in this sector and makes recommendations for on-going research and development that the geospatial community should focus on. It briefly examines recent progress in intersecting areas of technology affecting both framework data and LBS, such as service oriented architectures, quality assurance, mobile communications and near-real-time update. It then discusses developments in the establishment of persistent geospatial interoperability test-beds within the European research community and how these developments may help increase the productivity of geospatial research and make the research results more pertinent to commercial applications.

1. BACKGROUND

The concept of mobile location-based services for business and mass-market consumer applications already has a long history. The possibility of a mobile phone which:

- always “knows” where we are
- which, with their agreement, knows where our family, friends and colleagues are,
- that can search the Internet to identify information that we might need or be interested in, and
- which can then navigate us by foot, car or public transport efficiently and precisely to that location,
- which, en-route, provides us with up-to-the minute information such as details of traffic congestion, special offers from shops we pass etc.

is highly attractive and conceptually appealing.

The appeal of such services is in fact one of the reasons why in the late 1990's and early 2000's telecommunications companies paid billions of pounds in auctions for licences to run 3rd-generation mobile phone services (over £22 billion in the UK alone). Yet the services that have actually appeared have disappointed and, with the exception of the recent explosion of demand for in-car navigation systems have failed to meet user expectations.

The gap between concept and reality does not have a single cause but is a result of a shortfall in several areas. The ability to provide a mobile phone at a price which is acceptable to the mass-consumer that incorporates a high-quality GPS together

with all of the other facilities expected in today's modern phone has been a struggle that is only just being achieved. Even then there are problems such as battery capacity and interface complexity that have made such phones bulky or unappealing to a consumer whose desire for the capability offered is tempered by their desire also to have a sleek, slim and lightweight phone which is as much a fashion accessory as a box of electronic wizardry. The short-comings of global satellite positioning systems, principally the US GPS, as a tool for precise positioning of pedestrians in dense urban areas and in-side buildings has also meant that applications in this context have at best been incomplete and lacking viral appeal. The addition of inertial navigation sensors and the additional use of signals of convenience may help to address this problem though fully ubiquitous indoor and outdoor positioning is still a little way off. The quality of the location-based data, both the mapping and the geo-referenced “yellow-pages” data, has been a third area where the short-comings have meant that the appealing LBS concept has not matched the reality of what can be delivered. However, because of the technical limitations at the device end these data short-comings have been less apparent but are coming more to the fore as the other issues are one by one being resolved. This paper concerns itself with this last area and particularly the production and quality control of the framework map data.

* Corresponding author.

2. RECENT PROGRESS

2.1 The Impact of Mainstream IT and of Standards

The pattern of progress amongst the providers and maintainers of framework data, typically National Mapping and Cadastral Agencies (NMCA's) has been evolutionary in comparison with the dramatic and often disruptive developments in mobile phone devices. Understandably it has been data-centric, with a strong emphasis on data modelling techniques, as well as an increasing adoption of mainstream IT enterprise architectures. There has also been a growing concern, stemming from competitive pressures and the demands of users especially in critical applications, with data quality and currency. Overall Standards, both from mainstream IT and from the ISO TC211 and the Open Geospatial Consortium (OGC), are having a greater impact. But successful production environment implementations are emerging. Key elements at the system architecture level, covered in the rest of this section, are formal data modelling, Model Driven Architectures (MDA), schema translation, Service Oriented Architectures (SOA's) and formal workflow management. Underpinning advances in Knowledge Management techniques are covered in section 3.

2.2 Formal Data Modelling, UML and OCL.

Formal data modelling using Universal Markup Language (UML) has become well established in the geospatial domain in recent years. A UML class diagram describes the types of objects and the various kinds of structural relationships that exist between them, such as associations and subtypes. In addition the UML diagrams show the attributes and operations of a class and the constraints that apply to the way objects are connected. (Booth, Rumbaugh, Jacobson, 1999) Object Constraint Language (OCL) is a declarative language for describing rules that apply to UML and is now part of the UML standard. It is a precise text language that provides constraint and object query expressions on a model or metamodel that cannot otherwise be expressed by diagrammatic notation. There has been some success in using OCL within UML in a geospatial context, but workers have also reported limitations (Smits et al, 2005).

A recent EuroSDR Workshop on Feature/Object Data Models (EuroSDR, 2006) documented the emergence of a new generation of data models and their characteristics. The widespread use of UML for defining and documenting was clear. The reasons for this included better communication and definition at the conceptual level, the ability to manage evolution of the data model in a controlled manner, the emergence of suitable tools , including support for deriving traditional data model formulations such as Feature Catalogues from the UML model and the emergence of tools to automate the generation of implementation level schema from the conceptual level UML schema. The very considerable investment of effort in standards in this area, both in ICT and more specifically for the geospatial domain by ISO TC211 and the Open Geospatial Consortium (OGC), is now reaching the practical level with the emergence of powerful standards-based data models capable of supporting interoperability. Van Oosterveld and Lemmens (2006) describe the use of this approach in the generation of a 'core cadastral domain model':

'One of the advantages of modelling in UML is that it gives the possibility to generate an exchange format for the data in a standardised way. The GML3 standard (ISO 19136) describes

how to translate an UML model to a GML Applications Schema. This Application Schema uniquely defines an exchange format for data in the UML model. For the correct generation of such a schema the UML Model has to adhere to the encoding rules that are given in the GML Standard. Various tools exist that automatically convert an UML Model to a GML Application Schema. The ShapeChange tool (Portele, 2004) reads an UML Schema in the XMI exchange format and writes an XML Schema. The UML/INTERLIS Editor (Eisenhut, 2004) has an export button to generate a GML Application Schema.'

It is worth noting that the UML level model does not mandate whether the implementation should be centralised or decentralised.

2.3 Model-driven Architecture (MDA)

The essence of the MDA approach is that the conceptual (or platform independent <PIM>) model, usually expressed in UML, is in the driving seat. The storage (or platform specific, physical model <PSM>) is derived from it in an automated or near-automated fashion as is the mapping from the storage model to the transfer model (usually a GML Application Schema) as described in 2.3 above. Changes in the conceptual model can thus be propagated across the whole system in a co-ordinated and error free manner (see Figure 1).

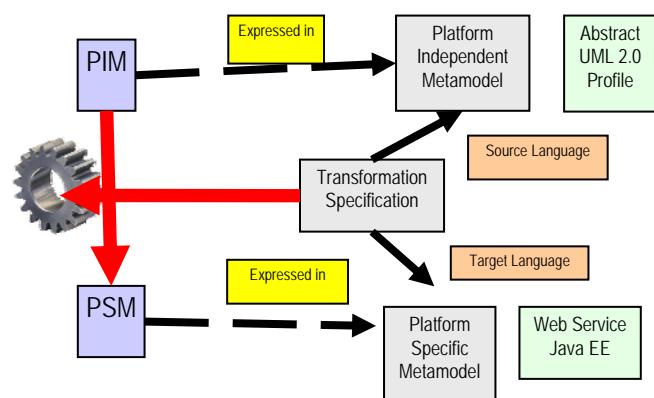


Figure 1. Change Propagation in MDA.

2.4 Schema Translation

Another key development has been that of schema-aware software components. XML and XMLSchema, together with domain specific extensions such as Geography Markup Language (GML) have gained near-universal acceptance and are supported by an impressive array of software tools. The role of the XMLSchema as a machine-readable manual for the data model underpins powerful schema translation techniques which in turn enable key functionality including:

- Elaboration, or publishing data to a variety of delivery schema from a source data model
- Data Harvesting, or the consolidation of diverse data sources to present according to a master schema, as required in Spatial Data Infrastructures (SDI's), (Snowflake, 2008)
- MDA's. The ability of system components to adapt automatically to schema changes is a fundamental to the realisation of the benefits of MDA.

2.5 Orchestration and Workflow Technologies

Advances in orchestration and workflow technologies have been standardised in the Web Services Business Process Execution Language OASIS Standard WS-BPEL 2.0, (OASIS, 2007, Wikipedia, 2008). BPEL is an XML grammar for ‘Programming in the Large’ and provides central orchestration of distributed components. Overall business process and workflow is implemented as a two tier system;

- Large scale process implementation – why do it? This level is governed using BPEL to control high level transitions in process (Abstract Process)
- Small Scale Business Logic implementation – what to do. This level is implemented by rules-processing engines, as described below in section 3.

BPEL decouples service implementation from process using WSDL/SOAP (W3C, 2003) messaging (see Figure 2).

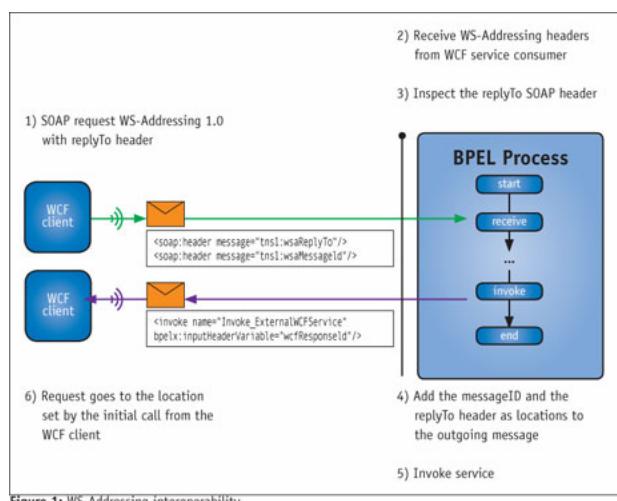


Figure 2. WSDL/SOAP Messaging

BPEL is widely implemented by mainstream IT software providers such as Oracle, IBM and Microsoft and is also available as Open Source. It is being used in the geospatial data management domain in the implementation of large scale management and update systems such as the use case described in section 4 below. In such an implementation the BPEL engine uses SOAP messaging to coordinate workflow across such activities as:

- Job Manager
- Job Manager Client
- Transaction Manager
- Feature Provisioning Service
- Feature Update Service
- Validation Manager

2.6 Service Oriented Architectures

OASIS defines Service Oriented Architectures (SOA) as the following:

A paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations.

The key elements of an SOA are illustrated diagrammatically in Figure 3. The SOA provides the overarching framework in which the aforementioned advances can be brought together, realising the following key characteristics:

- loose coupling
- formal contracts
- abstract logic
- reusable
- composable
- autonomous
- stateless
- discoverable
- encapsulation

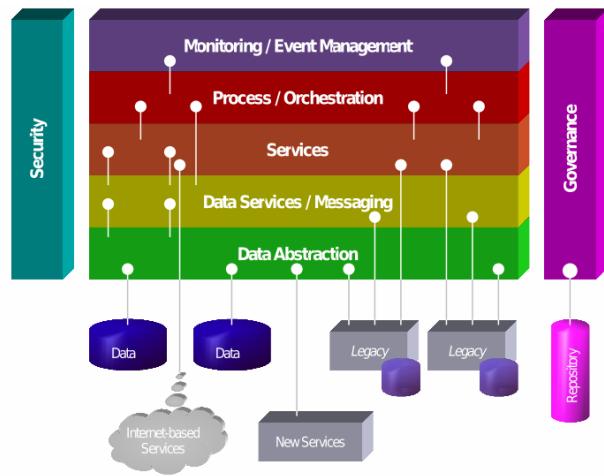


Figure 3 The Components of a SOA
(http://en.wikipedia.org/wiki/Image:SOA_Metamodel.svg)

3. RULES-BASED PROCESSING AND ADVANCES IN KNOWLEDGE MANAGEMENT

3.1 Rules and Rules Languages

Rapid advances in Knowledge Management in Information Technology are beginning to result in benefits in the geospatial realm. Historically, knowledge is buried inside data for example by the juxtaposition of data components or in implicit relationships. Alternatively it can be hidden in point applications or simply hidden inside people’s heads. This results in serious problems in keeping such knowledge up to date. Progress towards rigorous semantics contributes to removing ambiguities and to storing the knowledge/expertise of the organisation where everyone can contribute to it and share it, as enterprise metadata that is portable and independent of specific datasets and systems. A critical component in this development is a rules language to enable logical constraints to be specified. Such a language needs to be unambiguous, logical and portable, compact, intuitive, quantitative, web compatible, declarative and refinable. For a full discussion of these requirements and potential choices of rules languages, see (Watson, 2007). This area has received much attention of late through initiatives such as the Semantic Web community (W3C, 2004b) and rapid progress can be anticipated. There are currently several candidates to consider as a knowledge representation language (RDF, OWL, XML Rules/SWRL). None as yet cover all the functionality needed in the geospatial domain.

3.2 Rules-Based Processing

Rules-based processing follows the FACT-PATTERN-ACTION dynamic. Given some facts, if they meet any of the patterns/rules, perform the defined action(s). FACTs are a known data source. PATTERNs are the business rules that the data source obeys or should obey. ACTIONS happen as a result of PATTERNs being applied to FACTs as illustrated in Fig. 4. One application of rules-based processing is the implementation of the small scale business logic level of a workflow.

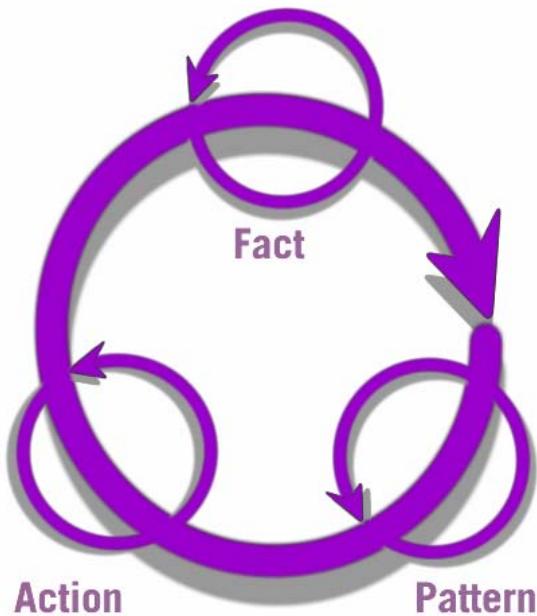


Fig. 4 The Fact-Pattern-Action Dynamic.

Of course, FACTS may not be 100% reliable and there is considerable scope for research into probabilistic association and the modelling of uncertainty. Also some subject domains are better defined than others and hence more amenable to rules-based processing.

3.3 A Rules-Based-Processing Environment

Rules-based processing is by no means new in GIS systems and has often been implemented using Object-Oriented, for example in the well known Smallworld GIS (Chance, Newell and Theriault, 1990). Radius Studio (1Spatial, 2008) from 1Spatial (formerly Laser-Scan) is an example of a modern rules-based processing environment, implemented both as middleware and as a service, that is used for business logic workflows and for domain discovery and conformance checking. A description of the role of rules-based processing in spatial data management and update is to be found in (Woodsford, 2007). Some key elements are addressed here, using Radius Studio as the example.

3.3.1 Defining and Validating the Rules-Base.

The rules-base is a set of conditions that objects from the data store should satisfy. A rule is a tree of predicates against which objects can be tested. Rules are expressed in a form independent of the schema of any particular data store. This means they can easily be re-used with different data sources.

The rules may be defined in text form, perhaps in conjunction with a logical data model or feature catalogue. They may be formally expressed in an ontology language such as Ontology

Web Language (OWL, W3C 2004a), in which case they can be directly used (by interfacing with the open source Jena ontology library (Jena, 2007)). More often the rules are not explicit or formalised, but exist in the form of knowledge held by domain experts.

An intuitive web-based interface is provided for defining rules and building up a rules-base (see Figure 5). The rules builder allows the definition of potentially complex rules with an easy to use, tree structured browser interface. The rule builder uses a syntax that is very much like English, so programming experience is not necessary. The validation of the rules-base by applying it to exhaustive representative datasets is a key stage before using it in earnest.

```

Rule condition for area objects:
  If ... Then ...
    area(area.geometry) is less than 2500
    Logical and:
      There are no woodland objects for which ...
        area.geometry is contained within woodland.geometry
      There are no water objects for which ...
        water.geometry contains area.geometry
  
```

Figure 5. The Rules-Definition Interface

3.3.2 Actions and Action Maps

An action is a procedure (or set of procedures) to be applied to one or more objects, usually when they are found to violate a particular rule. Actions are expressed in a form independent of the schema of any particular data store, so that they can easily be re-used with different sources of data. Actions are defined using a similar graphical user interface as for defining rules, but can also include operations such as assignment, conditionals and sequencing, object creation and deletion and report generation. Actions can be applied to all the objects from a data store, or in a more targeted manner by use of action maps.

An action map is an ordered list of (rule, action) pairs. When an action map is invoked, objects are checked against the rules. If the object does not conform to that rule, then the associated action is invoked. Action maps are often used to associate fixes to problems with non-conforming objects.

3.4 Rules-Based Processing and Data Quality

Radius Studio can be used to measure the quality of data in the sense of measuring the degree of conformance of the data to a rules-base. A document or data store can be processed and a report generated with results of the conformance test at both a summary and individual feature level. The summary reports the proportion of compliant objects. Additionally a list of non-compliant objects is generated (as HTML or as an XML-file,

with object identifiers). This consists of per-object metadata detailing which object infringed which rule. The rule identifier and text is included in each Object element with any unique key attributes and the bounding box (envelope) of the checked feature geometry attribute. Detailed feature level metadata contained within the Object elements can be used in manual reconciliation processes to locate and correct data.

Of course, the value of such a process as a measure of quality is dependent of the validity/quality of the rules-base. Quality can be improved by automatically applying ‘fixes’ to cases within allowable tolerances or parameters, as defined by an action map, or by referring the list of non-compliant objects to an interactive client. Tasks can be added, deleted and re-ordered. All required parameters can be browsed and updated. Pauses can be inserted at any stage in processing sequences and multi-level undo/redo can be used to correct mistakes. When an acceptable level of quality is reached, all objects that have been changed in the Radius Studio workspace are committed back to the main data store.

4. SOME USE CASES

4.1 A Very Large Data Provider (eg NMCA).

A typical case might involve a central master database with, say, 0.5 billion objects, and implementing thousands of changes per day by a combination of photogrammetric and field survey stations, both on site and detached (remote). A SOA provides the necessary degree of robustness, scalability and future-proofing. Use of the MDA approach supports evolution of the data model across the central data store and all the update processes. Centralised quality control and formalised workflow are also key factors in the implementation of such systems.

There may be sound reasons to have different data models for update and for data delivery, in which case schema translation can be used to keep the two data stores in synchronisation. Similarly the source data may be distributed across several data stores, with different data models and update regimes. ‘Intelligent piping’ to connect these to the master store involves on-the-fly schema translation.

4.2 Data Aggregator and Value Adder

A data aggregator and value added supplier for such applications as LBS and navigation has a very strong case for using rules-based processing for data conflation and for using a centralised rules-repository for quality control. Indeed such a repository will encapsulate much of the key knowhow of the enterprise and constitute important intellectual property. We can foresee the emergence of skilled practitioners of a new discipline – ‘Rules Authorship’ within the field of ‘Knowledge Management’.

5. THE ROLE OF TESTBEDS

The above text illustrates that substantial progress has been made over the last few years in the development of standards and techniques for data modelling, management and orchestration of workflow. However, the technology, if not intrinsically difficult to comprehend, does require an in-depth low-level understanding for implementation. Whilst the numbers of specialists needed with such skills may be small

there is a need for entrepreneurs and managers to understand what is realistically achievable in the context of operational services.

Furthermore, whilst not discussed in this paper, there have also been dramatic developments in technology for data capture and the breadth and volume of geospatial data potentially available to LBS has grown greatly. However, issues remain that need continuing research, for example, in the area of data generalisation or schematisation for intuitive communication of spatial data on small screen devices and in the area of semantic interoperability. The intrinsically distributed nature of the workflow and services to achieve compelling location-based services requires a collaborative approach at all levels; in the data acquisition and supply, in the creation of the component services and their delivery and equally in the research to address outstanding challenges.

To provide the necessary collaborative environment the Open Geospatial Consortium pioneered the use of community test-beds for rapid standards definition, development and testing. These test beds (see www.opengeospatial.org) have proved extremely effective and have led to a number of initiatives for the establishment of persistent test-beds that can serve specific communities.

In the area of research one good example is a tri-organisational initiative (EuroSDR, AGILE and OGC) launched in May 2007. This initiative (Gould et al, 2007) is for the development of a persistent geospatial interoperability test bed (PTB) to support the European academic and research communities. As of May 2008, four Use Cases have been defined to be provided by seven Universities from Germany, the Netherlands and UK. More than twenty other European academic and central and local governmental agencies have also expressed a desire to participate or support this initiative through the provision of data or services.

In the commercial world a parallel initiative has been launched led by Oracle Corporation and TeleAtlas known as the European Centre for Innovation in Geospatial and Location Based Services (www.innovation-geo-lbs.com). Again the development of a persistent test-bed is a key objective, in this instance to promulgate understanding of what can be achieved and to stimulate entrepreneurs to develop innovative location-based services that provide direct business benefits or to prototype applications of mass consumer potential.

6. CONCLUSIONS

The challenging demands on data management technologies driven by the LBS phenomenon around the turn of the last century were never fully exposed as a number of serious shortcomings at the device level, in positioning and in the user interface meant that large-scale adoption of LBS failed to materialise at that time. However, the inadequacies at the data and data management level were recognised by the system architects. This recognition led to renewed efforts in terms of appropriate data capture methods and research and development in the data management and quality control techniques and processes.

Steady and in some cases impressive progress has been made since then in resolving these issues. Key factors have been the maturing of more formalised approaches to data modelling, data

quality and workflow management, facilitated also by the move towards Service-Oriented Architectures. This progress is underpinned by Standards which are increasingly moving from the theoretical to practical implementation status assisted by test-bed initiatives involving community participation.

Advances in the individual topics dealt with here are in fact inter-related. More formalised data models underpin schema translation and the successful realisation of MDA and SOA. More formalism in the rules-based definition of data quality promotes the realisation of effective quality assurance alongside the other benefits of formal workflow management. There is encouraging evidence that a virtuous circle is emerging that will ultimately meet the most extreme challenges of data management posed by very large LBS and by national and international SDI's.

However, significant research challenges remain in a number of distinct areas. One example is semantic interoperability which is especially complex in the area of mass-consumer applications where formal scientific terminology does not apply and where usage and even the meaning of terms can change rapidly as services evolve. Issues of data fusion and the statistical and modelling implications of conflating many sources of data from different originations, collected by different sensorss and with different temporal, spatial and metadata quality characteristics is another area requiring continuing research. A third research theme relates to data representation, generalisation and schematisation for small screen devices. Progress of research in complex areas such as these requires the efficiency of the academic process to be addressed and linked to what may be considered to be more traditional commercial issues of scalability, performance, resilience and security. The development of standards and their application through collaborative mechanisms such as persistent interoperability test-beds is one way of accelerating and focussing research and linking the research to the associated development challenges which must be overcome before operational commercial use can become a practical reality.

REFERENCES

(Validated on 9 May 2008)

1Spatial, 2008. Radius Studio Home Page at:
http://www.1spatial.com/products/radius_studio/

Booch, G., J. Rumbaugh, and I. Jacobson, 1999, 'The Unified Modeling Language'. User Guide. Addison-Wesley Technology Series, Addison-Wesley, 1999.

Chance, A., Newell, R.G., Theriault, D.G. 1990. An Object-Oriented GIS – Issues and Solutions, in *EGIS '90 Conference Proceedings, Amsterdam, April 1990*.

Eisenhut, C. , 2004, Introduction to the UML-Editor. Available at: http://www.umleditor.org/download/refman_en.pdf

EuroSDR, 2006. Feature/Object Data Models Workshop in EuroSDR Official Publication No. 49. Available from www.eurosdr.net

Gould, M.; Jackson, M. J.; Klien, E. and Lemmens, R. 2007. Design and Implementation of a Persistent Testbed for Geo Web Services, 13th EC-GI & GIS Workshop, Porto, Portugal, 4-6 July 2007. Available at:

http://www.ec-gis.org/Workshops/13ec-gis/presentations/5_sdi_technology_III_Klien_2.pdf

Jena, 2007, <http://jena.sourceforge.net/>

Portele, C., 2004, Mapping UML to GML Application Schemas. Available at: <http://www.interactive-instruments.de>

OASIS, 2007. <http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf>

Smits, P. et al (2005). ESDI Conceptual Schema Workshop. Available at:
http://www.ec-gis.org/sdi/ws/csl/presentations/smits_summary_v5.ppt#1

Snowflake, 2008. Schema Translation for SDI (Spatial Data Infrastructure) – a White Paper available at:
<http://www.snowflakesoftware.com/news/papers/Schema%20Translation%20for%20SDI.pdf>

van Oosterom, P and Lemmen, C. 2006. The Core Cadastral Domain Model: A Tool for the Development of Distributed and Interoperable Cadastral Systems
Available at:
http://www.fig.net/commission7/india_2006/papers/ts04_01_lemmen.pdf

W3C, 2003. SOAP – Simple Object Access Protocol.
<http://www.w3.org/TR/soap/>

W3C, 2004a. Web Ontology Language (OWL).
<http://www.w3.org/TR/owl-features/>

W3C, 2004b. Semantic Web. <http://www.w3.org/2001/sw/>

Watson, P. 2007. Formal Languages for Expressing Spatial Data Constraints and Implications for Reporting of Quality Metadata. 5th International Symposium on Spatial Data Quality, ISSDQ 2007, Enschede, NL. ISPRS Working Group II / 7, available at:
[http://www.itc.nl/ISSDQ2007/proceedings/Session%205%20Dissemination%20and%20Fitness%20for%20Use/paper%20Paul_Watson\[1\].pdf](http://www.itc.nl/ISSDQ2007/proceedings/Session%205%20Dissemination%20and%20Fitness%20for%20Use/paper%20Paul_Watson[1].pdf)

Wikipedia, 2008. <http://en.wikipedia.org/wiki/BPEL>

Woodsford. P. A. 2007. Spatial Database Update and Validation – Current Advances in Theory and Technology. *ISPRS Archives Vol XXXVI, Part 4/W54, pp.1-7* available at:
<http://isprs-wg41.nsdi.gov.cn/Xinjiang/Proceedings/papers/1-7%20P.%20A.%20Woodsford%20.pdf>

ACKNOWLEDGEMENTS

The authors are pleased to acknowledge the contributions of many fellow workers including Paul Watson of 1Spatial, Eddie Curtis of Snowflake Software.