Linked Data in SDI
or How GML is not about Trees

Sven Schade and Simon Cox

European Commission - Joint Research Centre,
Institute for Environment and Sustainability, Ispra, Italy
{sven.schade, simon.cox}@jrc.ec.europa.eu

Abstract. There is increasing interest in shared geospatial information spaces. In the context of Spatial Data Infrastructures (SDI), heterogeneity of legacy systems and variety in existing standards are central challenges. Linked Data has been suggested as a potential solution in both cases. In this paper, we explain why the Linked Data approach provides no novelty to the geospatial community, as the principles are isomorphic to structures which were introduced to SDI standards ten years ago. However, while transformation from these to conventional Linked Data representation is trivial, key elements of the standards have not been widely implemented in the SDIs. Shared information spaces based on linked data can provide a motivation for implementation of the principles.

THE UNUSED POTENTIAL OF GEOSPATIAL DATA ENCODING

Geospatial information spaces at various geographic scales are becoming frequently requested, (INSPIRE, 2004; Weets, 2006). These must address expert users as well as the public, and both are stakeholders in the Shared Environmental Information System (SEIS)[1]. There cannot be a single architecture for realization, so a system-of-systems solution is inevitable. The Global Earth Observation System of Systems (GEOSS)[2] provides a large-scale example.

In respect to expert usage, and especially in the context of Spatial Data Infrastructure (SDI), many research projects have identified heterogeneity of legacy systems and variety of existing standards as central challenges (Probst, 2004; Liebermann, 2006; Lemmens 2006; Andrei, 2008). Today, standards coming from International Organization for Standardization (ISO)[3], Open Geospatial Consortium (OGC)[4], and the Infrastructure for Spatial Information in Europe (INSPIRE) (INSPIRE, 2003) are only partial harmonized and appear in different versions. Their interplay is hard to establish, because developments are dynamic and each participant to a system-of-systems uses a specific selection of standard specifications (Klien, 2009). A state of play analysis has been carried out recently in the EUROGEOSS project [5].

Linked Data has been suggested as a possible solution for creating shared information spaces (Bizer, 2008; Bizer, 2009). The notion of Linked Data refers to a best practice for exposing, sharing, and connecting resources in the (semantic) web (Bizer, 2008). It is based on (i) the use of Uniform Resource Identifiers (URIs) (Coates, 2001) as reference points, (ii) the Resource Description Frame-

---

[1] Web page available from: http://ec.europa.eu/environment/seis/, accessed 16th January 2010.

[2] Web page available from: http://www.earthobservations.org/, accessed 16th January 2010.

[3] Web page available from: http://www.iso.org, last accessed 16th January 2010.

[4] Web page available from: http://www.opengeospatial.org, last accessed 16th January 2010.

[5] Web page is available from: http://www.eurogeoss.eu, last accessed 16th January 2010.

work (RDF) (Lassila, 1999) as basic structure for any form of description, and (iii) content-negotiation (Holtman, 1998) to allow a client to specify an acceptable representation.

Standards in the geospatial information (GI) community are, by design, compatible with the Linked Data philosophy. The initial version of Geography Markup Language (GML) (OGC, 2000), an XML-based encoding standard for GI that was released ten years ago, supported:

Using interconnected pieces of data instead of single documents with complex internal structure.

The implementation of GML was strongly influenced by RDF. Version 1 had an explicit RDF/XML implementation binding, and subsequent versions (OGC, 2001; OGC, 2002; OGC, 2007) implement an object-property pattern and include XML linking (W3C, 2001) in a manner that is iso-morphic to RDF. Furthermore, the viability of GI applications with a cartographic interface serving as a Linked Data client is amply demonstrated by Google Maps (Granell, 2010, forthcoming). KML (OGC, 2007b) data includes links embedded in HTML (thus with informal semantics), which are shown in a pop-up balloons attached to points on the map (e.g. Figure 1).

Figure 1: Location of a hermitage near JRC in Google Maps.

Contemporary Linked Data only differs in its technological approach, and can be directly projected to SDI. Thus, combining GML with the concept of content-negotiation, we suggest that current SDIs can easily be put on the Linked Data track. This would overcome some of the challenges in integrat-ing SDI technologies into more heterogeneous systems-of-systems, and at the same time bridge the gap to the semantic web community (Janowicz, 2010, forthcoming; Cox, 2010, forthcoming; Cox, 2010b, forthcoming). In support of this, here we raise awareness about the flexibility of GML data encodings and clarify the relation to the Linked Data concept.

The remainder of the paper is structured as follows. A brief background on SDI standards and Linked Data is presented (section 2). In section 3, we project the Linked Data approach to classical SDIs, with examples. Section 4 provides a discussion on required changes in the context of SDI, be-fore we summarize and outline future work in section 5.

BACKGROUND ON GEOSPATIAL AND LINKED DATA

Standards for SDI and especially for the encoding of geospatial data have a long history. The Linked Data notion is relatively novel. Here we describe the central elements to be projected to each other.

Linked Data

The essential pillars of Linked Data are traditional web technologies and use of light-weight techniques for data model representation. The former depends on the use of Uniform Resource Identifiers (URIs) as reference points. A URI may be used to uniquely identify both data and non-data resources (Berners-Lee, 1998). Resolvers map a URI to the physical location of the resource, or in the case of non-data resources to a description.

Linked Data is usually implemented as common HTML (W3C, 1999) for human interfaces, plus Resource Description Framework (RDF) for links with machine processable semantics. RDF provides a structure for any form of description, and is the basis of the semantic web (Berners-Lee, 2001). RDF describes resources in the form of triples (subject-predicate-object) (Klyne, 2004). A basic typing mechanism for subjects, predicates and objects is available as RDF-Schema (RDF-S) (Brickley, 2004). RDF-S allows for extensions in order to specify domain-dependent subtypes, and thus allow for a domain vocabulary in its own namespace. RDF comes with different encodings, one of which RDF/XML. The key elements of RDF/XML for linking are 'rdf:about' (identifiers or anchors) and 'rdf:resource' (pointers or links). Resources become a set in which elements are connected with links. By these means, users can navigate between data like browsing through web pages. Generally, each piece of data contains link(s) to other data. However, leaf nodes or endpoints of the graph may make use of any other format, which may not support linking. Content-negotiation in HTTP allows client applications (like browsers) to negotiate various data representations (Holtman, 1998).

Although RDF is recommended for implementing the Linked Data, as a single global model for all data sources, other structured formats can support semantic linking, e.g. GML (introduced below).

Features in GML

Geography Markup Language (GML) is an XML representation for geospatial data, in particular when modeled as *features,* following the ISO/OGC reference model (OGC, 2008). Listing 1 provides a typical example. A feature is a discrete resource with a *type* defined in a *domain of discourse*. Any feature has a unique id, which is encoded using the attribute 'gml:identifier' (Listing 1, lines 02 and 03). In addition, a feature may have many properties, including spatial/geometric, simple attributes, and associations with other features (lines 04 to 36). The latter express logical or physical relationships between features. This aspect of the feature model maps easily onto the RDF meta-model. Furthermore, sub-classing is implemented in GML, using the XML Schema substitution group mechanism. However, sub-property relationships are not usually recorded[6], so GML does not generally support the full reasoning capability of RDF applications.

---

[6] The XML Schema rules for substitutability and type restriction interact in a way that makes sub-typing of property elements very hard to achieve in a schema-valid way, so GML generally recommends against even trying.

Listing 1: GML instance with nested structure, retrieved from
http://gsv-ws.dpi.vic.gov.au/test/EarthResourceML/1.1/wfs (simplified).

```
01 <er:MiningActivity gml:id="er.miningactivity.3612.3">
02  <gml:identifier>http://gsv-ws.dpi.vic.gov.au/test/EarthResourceML/MiningActivity/3612.3
03  </gml:identifier>
04  <er:activityDuration>
05   <gml:TimePeriod>
06    <gml:beginPosition>1867-01-01</gml:beginPosition>
07    <gml:endPosition>1872-12-31</gml:endPosition>
08   </gml:TimePeriod>
09  </er:activityDuration>
10  <er:oreProcessed uom="t">3</er:oreProcessed>
11  <er:associatedMine>
12   <er:Mine gml:id="er.mine.3611">
13    <gml:name codeSpace="http://www.dpi.vic.gov.au/earth-resources">Perseverance</gml:name>
14    <gml:name codeSpace="http://www.ietf.org/rfc/rfc2141">urn:cgi:feature:GSV:Mine:3611</gml:name>
15    <er:occurrence>
16     <er:MiningFeatureOccurrence gml:id="er.miningfeatureoccurrence.mine.3611">
17      <er:positionalAccuracy uom="m">100</er:positionalAccuracy>
18      <er:specification xlink:href="#er.mine.3611"/>
19      <er:location>
20       <gml:Point srsName="EPSG:4283">
21        <gml:pos>147.20258 -37.47436</gml:pos>
22       </gml:Point>
23      </er:location>
24     </er:MiningFeatureOccurrence>
25    </er:occurrence>
26   </er:Mine>
27  </er:associatedMine>
28  <er:producedMaterial>
29   <er:Product gml:id="er.product.361200.3">
30    <er:productName codeSpace="http://www.dpi.vic.gov.au/earth-resources">Gold</er:productName>
31    <er:sourceReference xlink:href="urn:ogc:def:nil:OGC::missing"/>
32    <er:grade uom=" g.t-1">10.4</er:grade>
33    <er:production uom=" oz_tr">1</er:production>
34    <er:sourceCommodity xlink:href="http://gsv-ws.dpi.vic.gov.au/t/EarthResourceML/Commodity/au"/>
35   </er:Product>
36  </er:producedMaterial>
37 </er:MiningActivity>
```

The listing represents a prototypical example for the current use of GML, where data elements are structured as trees (Figure 2). Note, however, GML augments the *tree*-based XML data model with links encoded as XML attributes using URIs, to allow a more general *graph* of feature associations to be encoded like RDF. Links may refer to nodes within the same document, or to external resources. GML also follows RDF/XML by implementing both features (resources) and properties as XML elements, where the name of the resource element gives the type of the resource, and the name of the property element implies the semantics of the relationship. This isomorphism between GML and Linked Data, especially RDF, will be discussed next.
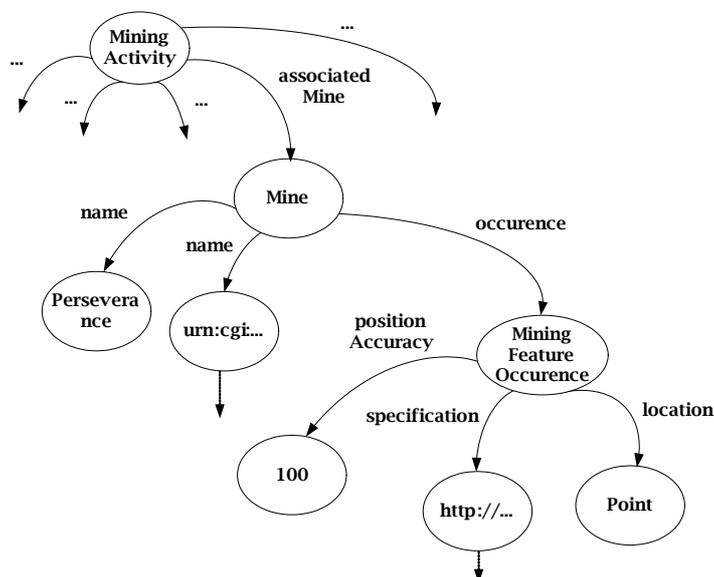
Figure 2: Structuring information about a mine as a tree.

THE (NEW?) WAY OF ENCODING GEOSPATIAL DATA

GML allows either in-line or by-reference representation of related resources, similar to RDF/XML. The W3C XML Schema formalization of GML also allows enforcement of either one of these encoding styles if desired. However, while 'rdf:about' and 'rdf:resource' were created specifically for RDF/XML, GML uses the standard W3C XLink mechanism ('xlink:href') to implement linking. Listings 1 and 2 show an example of a small dataset encoded in GML using nesting and linking, respectively. For example, the information about a mine and the produced material provided respectively in lines 11 to 17 and 28 to 36 in Listing 1 are replaced by links at lines 10 and 11 in Listing 2.

Listing 2: Normalized GML with associations implemented as links.

```
01 <er:MiningActivity gml:id="er.miningactivity.3612.3">
02   <gml:identifier>http://gsv-ws.dpi.vic.gov.au/test/EarthResourceML/MiningActivity/3612.3</gml:identifier>
03   <er:activityDuration>
04    <gml:TimePeriod>
05     <gml:beginPosition>1867-01-01</gml:beginPosition>
06     <gml:endPosition>1872-12-31</gml:endPosition>
07    </gml:TimePeriod>
08   </er:activityDuration>
09   <er:oreProcessed uom="t">3</er:oreProcessed>
10   <er:associatedMine xlink:href="http://gsv-ws.dpi.vic.gov.au/test/EarthResourceML/Mine/3611"
        xlink:role="http://www.earthresourceml.org/earthresourceml/1.1/xsd/mine.xsd#Mine"/>
11   <er:producedMaterial       xlink:href="http://gsv-ws.dpi.vic.gov.au/t/EarthResourceML/Product/3612.3"/>
12 </er:MiningActivity>
```

The basic mappings between GML and RDF are simple:

- xlink:href = rdf:resource
- gml:identifier = rdf:about

In Listing 3 the same example data is transformed to an RDF/XML representation. Lines 02 to 07 make use of the GML semantics of a time period, and in line 8, a GML data type is used, showing application of GML inside an RDF end point. The references to linked data about associated mines and produced materials (lines 09 and 10) are similar to the GML encoding presented in Listing 2. Each of these listings are supported by a schema which is not shown here, formalized in W3C XML Schema for Listings 1, 2, and in RDF-S for Listing 3.

Listing 3: RDF/XML representation.

```
01 <rdf:Description rdf:type="er:MiningActvity"
              rdf:about="http://gsv-ws.dpi.vic.gov.au/t/EarthResourceML/MiningActivity/3612.3">
02   <er:activityDuration
              rdf:dataType="http://schemas.opengis.net/gml/3.1.1/base/temporal.xsd#TimePeriodType">
03    <gml:TimePeriod>
04     <gml:beginPosition>1867-01-01</gml:beginPosition>
05     <gml:endPosition>1872-12-31</gml:endPosition>
06    </gml:TimePeriod>
07   </er:activityDuration>
08   <er:oreProcessed
              rdf:dataType="http://schemas.opengis.net/gml/3.1.1/base/basicTypes.xsd#MeasureType"
              uom=" t">3</er:oreProcessed>
09   <er:associatedMine
              rdf:resource="http://gsv-ws.dpi.vic.gov.au/t/EarthResourceML/Mine/3611"/>
10   <er:producedMaterial
              rdf:resource="http://gsv-ws.dpi.vic.gov.au/t/EarthResourceML/Product/3612.3"/>
11 </rdf:Description>
```

These representations are not merely isomorphic, but distinguishable only in minor aspects of syntax. The underlying information structure is the same: a directed graph (Figure 3). Thus GML can *in principle* be used directly for encoding linked data[7]. Furthermore, in GML Xlink attributes other than href may be used to attach additional semantics. For example 'xlink:role' may (optionally) indicate the type of the link target (e.g. Listing 2, line 10). In classic RDF the link must be traversed to discover this information.

---

[7] OGC has a legacy of using URNs to identify some resources, but has recently adopted an http URI scheme.
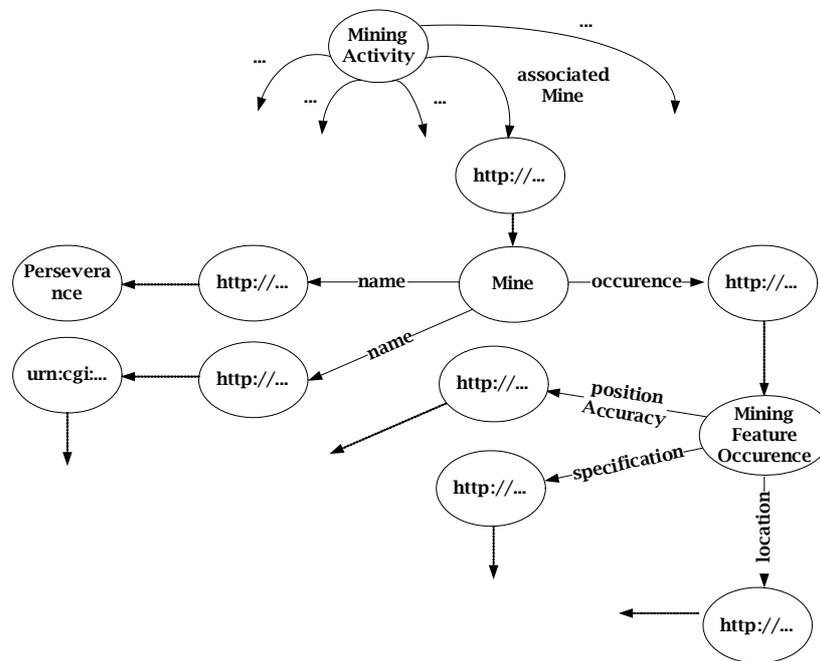
Figure 3: Structuring information about a mine as a (directed) graph (dashed lines represent links).

THE (NEW?) WAY OF PROVIDING GEOSPATIAL CONTENT

The examples in the previous section illustrate that:

1)   GML is easily transformed into a conventional RDF/XML, and
2)   native GML is essentially equivalent to RDF.

From these two facts we can conclude that XLink-aware GML clients are already prepared for Linked Data, with formal semantics for each link as defined by the GML application schema. However, this implies a requirement for providing content-negotiation in the GML context. Content-negotiation can be supported by providing different service interfaces operating on a shared data basis. This is where the technology intersects the SDI domain with the semantic web community, and we have two options for realization. On the one hand, we may only rely on semantic web interfaces, which are not standardized or harmonized in any way. On the other hand, we can move to a new generation of SDI service interfaces by changing/extending the established once. In favor of reuse, we propose following the second option. It would be easy to add a façade to a GML service (e.g. OGC Web Feature Service (WFS) (OGC, 2004) to transform the response into RDF, e.g. if a HTTP request header included 'accept=application/rdf+xml'.

Following a similar approach, Ordnance Survey (OS) of the UK has recently provided access to linked geospatial data[8]. As the underlying product (OS Master Map) is a GML application schema using the feature model and links between data elements were already established, so rich Linked Data has been provided quickly. The work focuses on reporting, data-centric scenarios still have to be specified. This work indicates a change in mindset and serves a test case, but illustrates feasibility on national level.

CONCLUSION AND FUTURE WORK

With this paper, we explicitly addressed the issue that SDI concepts and the notion of Linked Data do not exclude but complement each other. Linking (geospatial) data is a philosophy of usage and not a technical matter. Although the Linked Data concept has recently been applied to spatial data infrastructures, common and well established standards of the SDI community can be used for realization. In fact they were initially foreseen for doing so. In contrast to most XML, GML is not (and never was) about tree-like structures, but abut directed graphs. Due to historical reasons the initial RDF-based version was neglected in favor of less formal XML Schemas. Nevertheless, the basic structure remained. In this sense, GML could be characterized as 'premature Linked Data'.

The central requirement is availability of underlying structures, i.e. links, which can be provided on the fly or provided from storage for increased performance. Content-negotiation is the only core capability remaining to be implemented. We outlined how services (especially next generation WFS) can be developed for serving classical SDI and semantic web at the same time. We provided example data, and a similar approach could be taken for integrating data and metadata.

The decision of using a GML or RDF representation of geospatial data depends on the intended use. Both are optimized for different purposes. On the one hand, RDF allows for sophisticated querying and reasoning. Ontologies (Guarino 1998) can be utilized by particular RDF vocabularies, such as the Web Ontology Language OWL (Bechhofer, 2003) or the Web Service Modeling Language (WSML) (de Bruijn, 2006). On the other hand, numerous Geographic Information System (GIS) clients process GML.

Technical feasibility has been outlined and as initial implementations become available usage scenarios can be defined and analyzed. The assessment of benefits and development of prototypes as proof of concept is topic to future work. The EuroGEOSS project provides concrete test cases. Once added value has been proved, questions of responsible parties for link establishment and maintenance have to be identified and operational system development can proceed.

REFERENCES

Andrei, M., A.J. Berre et al., 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.

Bechhofer, S., F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, L.A. Stein, and F.W. Olin, OWL Web Ontology Language Reference, 2003.

---

[8] Web page is available from: http://data.ordnancesurvey.co.uk, last accessed 16th January 2010.

Becker,C. and C. Bizer, Exploring the Geospatial SemanticWeb with Dbpedia Mobile. Web Semantics: Science, Services and Agents on the World Wide Web, 2009.

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

Berners-Lee, T., J. Hendler and O. Lassila, The Semantic Web. Scientific American Magazine, 2001.

Bizer, C., T. Heath, and T. Berners-Lee, Linked Data: Principles and State of the Art. World Wide Web Conference, 2008.

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

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

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

Coates, T., D. Connolly et al., URIs, URLs, and URNs: Clarifications and Recommendations 1.0, 2001.

Cox, S. and S. Schade, Linked Data: What does it offer Earth Sciences? EGU General Assembly, 2010, forthcoming.

Cox, S., S. Schade, and C. Portele, Linked Data in SDI. INSPIRE Conference, 2010b, forthcoming.

de Bruijn, J., H. Lausen, A. Polleres, and D. Fensel, The Web Service Modelling Language WSML: An Overview. Berlin and Heidelberg, Germany, Springer, 2006.

Granell, C., S. Schade, and G. Hobona, Spatial Data Infrastructures and Linked Data. In Geospatial Web Services: Advances in Information Interoperability, IGI Global, P. Zhao (Editor), 2010, forthcoming.

Guarino, N., Formal Ontology and Information Systems. Formal Ontology in Information Systems. Amsterdam, Netherlands, IOS Press, 1998.

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

INSPIRE, Consultation Paper on a forthcoming EU Legal Initiative on Spatial Information for Community Policy-making and Implementation. 2003.

INSPIRE, INSPIRE scoping paper. 2004.

Janowicz, K., S. Schade, A. Bröring, C. Keßler, P. Maue and C. Stasch, Semantic Enablement for Spatial Data Infrastructures. Transactions in GIS, 2010, forthcoming.

Klien, E., A. Annoni, and P.G. Marchetti,. The GIGAS project – an action in support to GEOSS, INSPIRE, and GMES. In: J. Hrebícek, J. Hradec, E. Pelikán, O. Mírovský, W. Pillmann, I.Holoubek, T. Bandholtz (eds.): Proceedings of TOWARDS eENVIRONMENT: Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe, 2009.

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

Lassilia, O. and R. Swick, Resource Description Framework (RDF) Model and Syntax Specification, 1999.

Lemmens, R., A.Wytzisk, et al., Integrating semantic and syntactic descriptions to chain geographic services. IEEE Internet Computing, 10 (5), pp. 42-52, 2006.

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

OGC, OpenGIS® Geography Markup Language (GML) Encoding Standard - Version 1.0.0. The Open Geospatial Consortium, 2000.

OGC, OpenGIS Geography Markup Language (GML) Encoding Standard - Version 2.0. The Open Geospatial Consortium, 2001.

OGC, OpenGIS Geography Markup Language (GML) Encoding Standard - Version 3.0. The Open Geospatial Consortium, 2002.

OGC, OpenGIS Web Feature Service (WFS) Implementation Specification – Version 1.1.0. The Open Geospatial Consortium, 2004.

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

OGC, OGC KML - Version 2.2.0. The Open Geospatial Consortium, 2007b.

OGC, OGC Reference Model (ORM) - Version 2.0. The Open Geospatial Consortium, 2008.

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

W3C, XML Linking Language (XLink) Version 1.0. W3C Recommendation 27 June 2001.

W3C, HTML 4.01 Specification W3C Recommendation 24 December 1999.

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