Towards Meaningful URIs for Linked Sensor Data

Krzysztof Janowicz[1], Arne Bröring[2], Christoph Stasch[3], & Thomas Everding[3]

[1] Department of Geography, The Pennsylvania State University, USA
[2] 52° North Initiative for Geospatial Open Source Software, Germany
[3] Institute for Geoinformatics, University of Muenster , Germany

Abstract. Sensor data is stored and published using OGC's Observation & Measurement specifications as underlying data model. With the advent of volunteered geographic information and the Semantic Sensor Web, work on an ontological, i.e. conceptual, model gains importance within the Sensor Web Enablement community. In contrast to a data model, an ontological approach abstracts from implementation details by focusing on modeling the real world from the perspective of a particular domain or application and, hence, restricts the interpretation of the used terminology towards their intended meaning. The shift to linked sensor data, however, requires yet another perspective. Two challenges have to be addressed, (i) how to refer to changing and frequently updated data sets such as stored in Sensor Observation Services using Uniform Resource Identifiers, and (ii) how to establish meaningful links between those data sets, i.e., observations, sensors, features of interest, observed properties, and further participants in the measurement process. In this short paper we focus on the problem of assigning meaningful URIs.

1 Motivation

The Sensor Web Enablement (SWE) initiative of the Open Geospatial Consortium (OGC) is responsible for the development of standards to make sensors and their gathered data accessible on the Web. The Observation & Measurement (O&M) and Sensor Model Language (SensorML) specifications define how to exchange data about sensors and their observations, while a number of Web Services is responsible for their storage, retrieval, tasking, and access. For instance, the Sensor Observation Service (SOS) stores and gives pull-based access to observation data. Rigid standardization and conformance tests ensure that these services can be combined to Spatial Data Infrastructures (SDI) to support complex tasks and semi-automatic service chaining. With respect to the vision of a Digital Earth, such infrastructures do not only deliver data but also perform processing and simulation steps, as well as the final rendering on a digital map or virtual globe. Grounded in spatial and temporal reference systems, the results from different services can be integrated into a multi-layered representation of the earth and help to answer scientific questions or assist in emergency situations.

As there is no context-free and canonical representation of geographic features or even their corresponding types [1,2,3], a meaningful layering also requires the integration between the thematic components of geographic information, e.g., by grounding them in semantic reference systems [4,5]. This challenge becomes even more pressing taking volunteered geographic information into account in which citizens participate as sensors by creating and updating content for the Digital Earth [6]. Each information community has its own requirements and motivations for contributing data which is also reflected by differences in the conceptualization of geographic space and leads to semantic heterogeneity [7]. The Semantic Web addresses these challenges in two steps, (i) by providing formal and machine-readable specifications for the conceptualizations underlying these information communities, i.e., by creating ontologies, and (ii) by inference engines used to discover implicit relations, facts, and potential contradictions. Unfortunately, Spatial Data Infrastructures and the Semantic Web are mostly isolated from each other – a semantic integration layer connecting ontologies and reasoners with SDI services is missing.

In previous work, we have defined and partially implemented[1] such a Semantic Enablement Layer (SEL) [8]. It encapsulates Semantic Web reasoners and repositories by existing OGC services such as Web Processing Services (WPS) and, hence, enables a transparent and seamless integration of Semantic Web technologies into Spatial Data Infrastructures. Currently, we also focus on the reverse direction – namely to make spatial information available on the Semantic Web without changing existing SDI standards and implementations. An ongoing part of this work is the development of a RESTful SOS proxy for Linked Data on the Semantic Web. The proxy can be installed as a facade to a Sensor Observation Service without any modifications to the service interface or database and provides an RDF representation of O&M data, as well as URIs.

2 Towards a Linked Data Model for Sensor Data

Meaningful URIs, links, and vocabularies are the essential building blocks of Linked (Open) Data, while the machine-readable RDF is rather an enabling encoding that provides some basic reasoning support in addition. While transforming GML-encoded geographic information to RDF is an essential step in making it accessible on the Linked Data Web, a purely automated XSLT-based mapping would be of questionable value and does not add sufficient semantics [9,10]. A Linked Data model has to address the problem of assigning meaningful URIs to particular data chunks, the semantic annotation of data using ontologies, as well as establish pre-defined links between data. Each of these requirements has to be addressed, however, for lack of space this short paper focuses on the URI scheme. As sources of reference, URIs are of special importance as Linked Data detaches information from its original context, e.g., a document, application, or database. While this eases accessibility and supports re-usability it makes the interpretation of such data more difficult [7].

---

[1] For a first release of the Web Reasoning Service (WRS) as part of the SEL see `https://svn.52north.org/svn/semantics/WRS`. For the ongoing implementation of the RESTful SOS proxy visit `http://52north.org/RESTFul_SOS`.

2.1 A URI Scheme for Linked Sensor Data

To connect observations provided by Sensor Observation Services to the Linked Data Web requires URIs for the different components of the O&M model [11]. Going beyond the initial proposal of Sequeda and Corcho [11], the 52°N RESTful SOS proxy resolves such well-defined URI scheme and generates the according *GetObservation* calls to query the underlying SOS. The results provided for these URIs are RDF encodings of the O&M data.

The main O&M components associated with an observation and offered by an SOS are *features of interest*, *sensors* (procedures), and *observed properties*. The URIs are assigned to these components by appending the component type to the URI identifying the authority. For example, `http://my.authority.org/sensors` returns links to all sensor descriptions (in SensorML). Consequently, `http://my.authority.org/sensors/thermometer1` returns the description of a certain thermometer and links to the produced observations. An according scheme is defined for features of interest and observed properties.

For enabling the RESTful access to sensor observations, the base URI scheme has the form: `http://my.authority.org/sos/observations`. By following this base link all observations of the service are returned. Observations measured by specific sensors, gathered for particular observed properties, or from specific features can be accessed by appending identifiers of those resources to the base URI. For example, the reference `http://my.authority.org/sos/observations/thermometer1/RiverSegment23/temperature`, points to all observations gathered by the sensor *thermometer1* at the feature of interest *RiverSegment23* for the observed property *temperature*.

A reference to all temperature observations, for example, appends the segment `/-/-/temperature` to the base URI. The order of the sensor, feature, and observed property identifiers has to be preserved when constructing URIs. This ensures that the RESTful SOS proxy can determine the meaning of each URI segment and construct an according GetObservation requests. For this reason, the dash ('-') serves as wildcard.

Such URIs for observations are used to provide links from sensor and feature descriptions to their related observations. For example, the sensor description at `http://my.authority.org/sensors/thermometer1` contains links to the observations produced by the sensor: `http://my.authority.org/observations/thermometer1/RiverSegment23/temperature`. This ability to collect observations by following links offered by the RESTful SOS proxy replaces the extensive filtering capabilities of the original SOS interface.

To refer to observations from a particular time instant or period, additional tokens can be appended to the URI according to the scheme '<start date>/<end date>' where each date has the form '<year>/<month>/<day>/<hour>/<minute>/<second>'. This scheme is based on ISO 8601 [12], however, we use forward slashes as token separators. Fractions of seconds can also be expressed as defined by ISO 8601. Here the separator character is a dot; as a comma may cause problems and would have to be escaped. Similarly to the examples above, particular tokens can be replaced by wildcard character ('-'). For example, appending '/2010/02/-/-/-/-/2010/07/01/13/05/-' to an URI refers to all observations measured between beginning of February 2010 and 1st of July 2010 at 13:05.

After defining the temporal filter, a bounding box can be appended as additional spatial filter to the URI. The encoding of the bounding box is a list of underscore separated values; again a separator is chosen that does not require escaping and improves readability. The first four values are the coordinates defining a two dimensional rectangle, while the fifth value is the identifier of their reference system: <minCoord1>_<minCoord2>_<maxCoord1>_<maxCoord2> _<crsURI>. Since observations do not necessarily have a location property, the spatial filter is applied to the position of the feature of interest associated with an observation[2]. In contrast to the temporal filter, the coordinates are not in different URI segments as leaving one of them aside would be meaningless. For example, the URI segment for a bounding box may be specified as follows: '89000_834000_285000_962000_urn:ogc:def:crs:EPSG:6.5:4326'.

Finally, using geographic coordinates for the construction of URIs for RESTful access also leads to the question how to encode geometry in the Linked Data model. The geometry of features of interest can be either encoded as plain text coordinates list such as defined in GML or also transformed to a RDF representation of single points connected using some relation. The latter solution may lead to overhead as retrieving the geometry of a feature may involve traversing all points defined in the RDF encoding. This is only a feasible approach if the introduction of a labeled and directed graph adds any further reasoning or retrieval capability. So far, existing solutions do not require RDF serializations, e.g., for computing topological relations. This is another example demonstrating the difference between a conceptual model in which, for instance, the separation between latitude and longitude is meaningful, while it may be cumbersome for a Linked Data model.

2.2 How to Establish Meaningful URIs?

One postulation of the Linked Data Web is to make *raw data* available on the Web and assign Uniform Resource Identifiers to it. While this may seem to be straightforward, it leads to various problems including object identity [7], granularity, or allowed processing steps. OGC's notion of the *feature of interest* is used here for demonstration purpose. Observation is the act in which sensors perceive stimuli and translate them into another, often digital, representation. Typically, we are not interested in these stimuli but in what they tell us about the properties of particular real word entities - the features of interest.

However, there is no a priori conceptualization of geographic space and the creation of features and types is rather an act of cognition and social convention. In case of geographic information, the extraction of features requires several processing steps which are arbitrary to a certain degree. For instance, the extraction

---

[2] The SOS specification uses the 'value reference' property to identify the property to which the spatial filter is applied. For the sake of simplicity we restrict this property to the position of the feature.

of features from raster data depends on the used algorithms and application specific thresholds. The classical downtown problem can be used to illustrate this process [13]. There is no fixed region for, say, the *downtown* of State College, PA. One may extract such a vague regions from human participants tests, by investigating tags used by Web 2.0 photo communities [14], and various other approaches. Additionally, each of these methods may depend on particular confidence values or other parameters. The following example shows a URI for State College downtown extracted using a yolk-egg model based on a 50% confidence.

`http://my.authority.org/FOI/statecollege_downtown/yolk-egg/C50/`

Meaningful URIs also require a careful sequencing of the segments forming such a URI. For instance, leaving the *C50* aside should still refer to a resource – in this case, return all resulting polygons based on the yolk-egg model available for downtown State College. By further reducing the URI one receives all potential downtown representations extracted using different approaches from the data stored at my.authority.org. One could also argue that this is the base URI identifying the downtown of State College and providing all RDF-encoded information about the feature as well as links to the different geometries. While this is a feasible solution, it does not solve the related problems of identity as this would require a global, context-free notion of State College or complex semantic mappings. It is worth mentioning that using *owl:sameAs* between different versions of downtown would rather add to then resolve such confusions.

Finally, to put more focus on the notion of *raw* data, one could assign URIs to unprocessed, direct outputs of sensors. However, this is of questionable value for two reasons: (i) What is the right granularity for such data chunks? For instance, satellite data is recorded based on the *swath width* which clearly has no reference to any geographic features. Hence, assigning URIs to such huge data chunks would render them meaningless and will fragment features randomly over different data chunks. The opposite (and also meaningless) alternative would be to assign URIs to single pixels. (ii) One reason for the limited re-usability of sensor data is that by just deploying sensors we make various assumptions about the studied phenomena. There is hardly any context-free sensor data.

3 Conclusions and Outlook

In this work, we briefly introduced a URI scheme for a RESTful SOS interface for Linked Data. We also discussed the problem of identity for URI assignment which stems from an entity-centric view in contrast to the continuous fields of sensor observations. Further work will target the development of a Linked Data model as sketched in section 2. This especially also includes links between the sensor observations and other Linked Data. For reasons of simplicity, we have also not discussed the distinction between URIs referring to real world features as proposed by the *Web of Things* in contrast to URIs referring to data about these features. In the future, more SWE standards and specifications shall be included in to the Linked Data model - SensorML may be the most prominent candidate.

Acknowledgments

The presented work is developed within the 52°*North* semantics community. We are thankful for discussions with members of the Münster Semantic Interoperability Lab (*MUSIL*) and the US Geological Survey (*USGS*).

References

1. Mark, D.M.: Toward a theoretical framework for geographic entity types. In: Spatial Information Theory: A Theoretical Basis for GIS, International Conference COSIT '93. (1993) 270–283
2. Brodaric, B., Gahegan, M.: Experiments to Examine the Situated Nature of Geoscientific Concepts. Spatial Cognition and Computation **7**(1) (2007) 61–95
3. Frank, A.U.: Multi-cultural aspects of spatial knowledge. In Janowicz, K., Raubal, M., Levashkin, S., eds.: GeoSpatial Semantics, Third International Conference, GeoS 2009. Volume 5892 of LNCS., Springer (2009) 1–8
4. Kuhn, W.: Semantic reference systems (guest editorial). International Journal of Geographical Information Science **17**(5) (2003) 405–409
5. Scheider, S., Janowicz, K., Kuhn, W.: Grounding geographic categories in the meaningful environment. In Hornsby, K., Claramunt, C., Denis, M., Ligozat, G., eds.: Conference on Spatial Information Theory (COSIT 2009). Volume 5756 of LNCS., Springer (2009) 69–87
6. Goodchild, M.: Citizens as sensors: the world of volunteered geography. GeoJournal **69**(4) (2007) 211–221
7. Janowicz, K.: The role of space and time for knowledge organization on the semantic web. Semantic Web - Interoperability, Usability, Applicability (2010; to appear)
8. Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maue, P., Stasch, C.: Semantic enablement for spatial data infrastructures. Transactions in GIS **14**(2) (2010) 111 – 129
9. Jain, P., Hitzler, P., Yeh, P.Z., Verma, K., Sheth, A.P.: Linked Data is Merely More Data. In: AAAI Spring Symposium 'Linked Data Meets Artificial Intelligence', AAAI Press (2010) 82–86
10. Schade, S., Cox, S.: Linked data in sdi or how gml is not about trees. In: Proceedings of the 13th AGILE International Conference on Geographic Information Science - Geospatial Thinking. (2010)
11. Sequeda, J., Corcho, O.: Linked stream data: A position paper. In: Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09). Volume 522., CEUR-WS (2009) 148–157
12. ISO: 8601:2004(E) Data elements and interchange formats - Information interchange - Representation of dates and times (12 2004)
13. Montello, D.R., Goodchild, M.F., Gottsegen, J., Fohl, P.: Where's downtown?: Behavioral methods for determining referents of vague spatial queries. Spatial Cognition & Computation **3**(2) (2003) 185–204
14. Keßler, C., Maué, P., Heuer, J.T., Bartoschek, T.: Bottom-up gazetteers: Learning from the implicit semantics of geotags. In Janowicz, K., Raubal, M., Levashkin, S., eds.: GeoSpatial Semantics, Third International Conference, GeoS 2009. Volume 5892 of LNCS., Springer (2009) 83–102