Aggregating Linked Sensor Data

Christoph Stasch[1], Sven Schade[2], Alejandro Llaves[1], Krzysztof Janowicz[3], and Arne Bröring[145]

[1] Institute for Geoinformatics, University of Münster, Germany
[2] Institute for Environment and Sustainability, Joint Research Centre, Ispra, Italy
[3] University of California, Santa Barbara, USA
[4] ITC Faculty, University of Twente, Enschede, Netherlands
[5] 52°North Initiative for Geospatial Open Source Software GmbH, Germany

Abstract. Sensor observations are usually offered in relation to a specific purpose, e.g., for reporting fine dust emissions, following strict procedures, and spatio-temporal scales. Consequently, the huge amount of data gathered by today's public and private sensor networks is most often not reused outside of its initial creation context. Fostering the reusability of observations and derived applications calls for (i) spatial, temporal, and thematic aggregation of measured values, and (ii) easy integration mechanisms with external data sources. In this paper, we investigate how work on sensor observation aggregation can be incorporated into a Linked Data framework focusing on external linkage as well as provenance information. We show that Linked Data adds new aspects to the aggregation problem, e.g., whether external links from one of the original observations can be preserved for the aggregate. The Stimulus-Sensor-Observation (SSO) ontology design pattern is extended by classes and relations necessary to model the aggregation of sensor observations.

Keywords: Sensor Aggregation, Semantic Enablement, Linked Data

1 Introduction

Sensor observations are collected with a specific purpose in mind and, therefore, measuring follows strict procedures and spatio-temporal scales [1]. While the same device, e.g., a thermometer, can be used to measure soil and air temperature, both follow different procedures and their results cannot be combined. Similar issues hold for fine dust $(PM10)^6$ measurements, where data coming from rural monitoring stations has to be distinguished from data produced by sensors located in urban areas, particularly at major roads [2]. Consequently, the rich observation data gathered by today's public and private sensor networks is difficult to reuse aside of the initially intended context. We hope to boost the use of observation results and the number of innovative observation-based applications by providing mechanisms for (i) spatial, temporal, and thematic aggregation of measured values, and (ii) easy integration mechanisms with other data sources.

[6] The notation PM10 is used to describe fine dust particles of 10 micrometers or less.

Building up on our previous work on exposing standardized observation data as Linked Data [3], this paper introduces the next steps towards opening up sensor observations to new usage scenarios: the aggregation of observations and exposing them as Linked Sensor Data. Having temporal aggregates (e.g., yearly averaged fine dust measures), spatial aggregates (e.g., fine dust concentration in the Münsterland region in Germany), thematically aggregated observations (e.g., blizzards, landslides, or forest fires), and their combinations available, makes linking more attractive and opens environmental information to new user communities. On the one hand, observations may be connected to particular features of interest in the Linked Data cloud. On the other hand, hubs such as DBpedia may directly refer to aggregated observations, e.g., an entry about the German city of Münster and its surrounding areas by referring to recent and average weather conditions, or air quality measures.

The main contributions of this paper are threefold. We (i) present a Linked Data model for aggregated sensor data, (ii) discuss the effects of aggregation on links from and to observations, and (iii) outline the role of provenance in this setting. The implementation of the extensions discussed in this paper are ongoing and the 52°North semantics community[7] plans to release an updated prototype in fall 2011.

The remainder of this paper is structured as follows. In section 2.1, we introduce the concept of aggregated observations and provide background information about Linked Sensor Data and provenance information in observations. Section 3 discusses the implication of aggregation on Linked Sensor Data. Here, we present the required extensions to our Linked Data model for observations. Additional investigations address the effects of aggregation on external linking, and issues on data provenance. In section 4, we set our work in relation to current efforts to provide observations as Linked Data and to provide provenance information in observation data. The paper concludes with a summary and an outline of the remaining steps for implementing aggregated observations as Linked Data; see section 5.

2 Background

In this section, we provide a brief overview on related work. At first, we introduce definitions and related work about aggregation of observations. Second, we introduce the concepts of Linked Data. Finally, we describe related work about provenance of sensor data.

2.1 Aggregation of Observations

Aggregation of observations in space and time is essential to derive information that is useful for a certain application purpose and to integrate observation data with differing spatio-temporal resolutions. Yet, spatio-temporal aggregation of

[7] Implementations and documentations can be found at `http://52north.org`.

observations in the Linked Data context has not yet been addressed. However, in other communities, e.g., the database community or in environmental sciences, spatio-temporal aggregation has been a research topic for years and is sometimes also referred to as scaling of observations and environmental models. Vega and Lopez [4] give a comprehensive survey on spatio-temporal aggregation methods for databases. Besides simple aggregation, complex statistical models might also be applied as described for the domain of soil sciences by Bierkens et al. [5]. Spatio-temporal aggregation processes for observation data are not yet available on the Web and, therefore, recent approaches demonstrate how to tackle this challenge. A spatio-temporal aggregation service that can be used to provide such aggregation functionality on the Sensor Web has been introduced in our previous work [6].

In this paper, we largely follow the definition of aggregation[8] by Jeong et. al. [7]. During an *aggregation process*, the observations are grouped by a *grouping predicate*, e.g., by a spatial predicate which is defined by the polygon representing the area of a city, or by a temporal predicate defined by the time period of a month. After grouping, an *aggregation function* is applied that computes a single value, an *aggregate*, for the result values of an observation group. The aggregation function might be linear (e.g., MEAN), but also non-linear (e.g., MEDIAN, or areal fraction of spatial blocks where the concentration of a pollutant exceeds a critical level) [8]. The grouping predicate does not necessarily have to be the target spatial or temporal extent of an aggregated observation. Considering the example of a block kriging method [9], for every aggregate of a spatial block, all measurements are taken into account and not just the ones laying in the extent of the block. Similarly, for temporal aggregation, moving windows might be used to aggregate values to time periods that also include the values before and after a certain period.

Besides spatio-temporal aggregation as introduced above, extracting high level events from observations is also done by aggregating observations. Treating the high level events as observations again enables an easy integration into existing infrastructures and tools. Considering the blizzard example as described in [10], the event of a blizzard can also be modeled as an observation. The blizzard is an aggregate of several observations indicating heavy snowfall, very low temperatures, and high wind speed. This example demonstrates that the grouping predicate is not merely spatial or temporal, but also contains predicates on the result values of the observations (e.g., heavy snowfall). We thus refer to this kind of aggregation as *thematic aggregation* of observations. However, the observations are still aggregated spatially or temporally as the blizzard is observed at a region in space and for a period in time.

[8] Aggregation might be also referred to as complex entity with parts. In case of observations, this might be a collection of observations where the non-aggregated observations are parts of the aggregated observation collection. However, in our work we consider aggregation as described in this paragraph and commonly used in environmental sciences.

2.2 Linked Data

For aggregation of observations a mechanism that helps to retrace the original observations and sensors from the aggregated observations is important. Linked Data [11] provides a promising paradigm to provide such a mechanism, as the original observations and the aggregates can be easily linked with clear semantics. Linked Data proposes unique identifiers for data in the Web, links between them, and relies on the Resource Description Framework (RDF) [12]. The most common query language for RDF is SPARQL [13]. SPARQL has similar capabilities as query languages for relational databases, but works by matching graph patterns and is optimized for RDF triple stores, such as Sesame or Virtuoso. Within the last years, Linked Data has become the most promising vision for the Future Internet and has been widely adopted by academia and industry.

Several approaches for Linked Sensor Data in the Web are already available [14,15,16]. They describe, how to identify sensor resources using URIs, how to link them with clear semantics and how to expose the sensor data in the Web. However, the issue of spatio-temporal aggregation, e.g. how aggregation affects the links from and to observations, is not yet addressed. In our previous work [3], we developed a standards-based approach to expose sensor metadata and observations stored in a Sensor Observation Service (SOS) [17] to the Semantic Web by following Linked Data principles and providing dereference-able HTTP URIs for sensors, observed properties, features of interest, and observations, link them (to external sources), and expose their semantics using the SSO ontology [18]. In this work, we extend our previous work on Linked Sensor Data to support aggregated observations.

2.3 Provenance in Observation Data

There are several approaches available for providing provenance information in the Web. The W3C's Provenance Incubator Group[9], predecessor of the new Provenance Working Group[10], compiled a list of requirements to support provenance in RDF, which includes for example that every observation should have an URI identifier [19]. Based on these requirements, the Provenance Vocabulary has been defined[11] that can be used in the Web to provide provenance information for Linked Data [20]. Similar to the Provenance Vocabulary, the Open Provenance Model[12] (OPM) defines nodes and edges to create provenance graphs that allow to retrace the creation of an item back to its origin. The nodes can be *artifacts*, *processes* and *agents* whereas the edges between nodes can be defined as the causal relationships *used*, *wasGeneratedBy*, *wasControlledBy*, *wasTriggeredBy*, and *wasDerivedFrom*. The graphs can be serialized in different data formats like XML.

[9] http://www.w3.org/2005/Incubator/prov/charter
[10] http://www.w3.org/2011/01/prov-wg-charter
[11] http://sourceforge.net/apps/mediawiki/trdf/index.php?title=Provenance_Vocabulary
[12] http://openprovenance.org/

Besides general approaches for provenance information in the Web, providing provenance information in Linked Sensor Data has recently gained attention. Provenance of sensor data can be defined as information about the source of the sensor data as well as information about transformations applied to the original data [21]. Patni et al. [10] propose an approach for provenance in Linked Sensor Data and define the capabilities of the sensor, the spatio-temporal parameters of the observation, and the measurement value as relevant sensor provenance information. Liu et al. [22] introduce a provenance aware virtual sensor system based upon the OPM. Using the OPM for their virtual sensors enables the description of (i) fetching processes for sensor data streams; (ii) workflow execution like data transformation of raw measurements; and (iii) user interaction with a web application that allows to manage the virtual sensors. In another approach, Park and Heidemann [21] defined their own provenance model that (for sensor data) is more comprehensive than the OPM. Among other things, this alternative model allows the definition of access control for sources. Similar to the approach of Liu and colleagues, the sensor data is annotated with additional provenance metadata. Our approach will show how most of relevant provenance information is already provided in our Linked Sensor Data and how the links can be mapped to provenance relationships as defined in the OPM.

3 Aggregation of Observations in the Linked Data Cloud

In this section we introduce an approach to enable the aggregation of observations in the Linked Data cloud. First, we present an extension of the Stimulus-Sensor-Observation (SSO) ontology design pattern [18]. Next, we illustrate how the change of observation properties during aggregation affects the links from and to observations in the cloud. Finally, we describe how provenance information pointing back to the original observations can be provided.

3.1 Extension of the SSO Design Pattern

Following our previous work [3], we use an intermediate Linked Data model for exposing sensor observations. It was derived from an ontology developed by the W3C SSN-XG [23], namely the Stimulus-Sensor-Observation (SSO) ontology design pattern [18]. The SSO pattern forms a generic and adaptable starting point for the development of sensor ontologies as well as Linked Data vocabularies.

Figure 1 shows the classes and relations from the pattern extended by the Linked Data model for sensor data, and the new elements that have been added in order to account for aggregation. In a nutshell, we reuse the following definitions:

- *FeatureOfInterest*: entity that comprises observable properties.
- *ObservedProperty*: property that inheres in a feature of interest.
- *ObservationCollection*: set of observations, grouped by a distinct criteria.
- *Observation*: (social) construct that connects observed properties with sensors, sensing results, and sampling times.

– *SamplingTime*: time instant or interval at which an observation was made.
– *Result*: symbol representing an observed value.
– *Sensor*: entity that performs observations.
– *Procedure*: description that specifies how observations have to be carried out

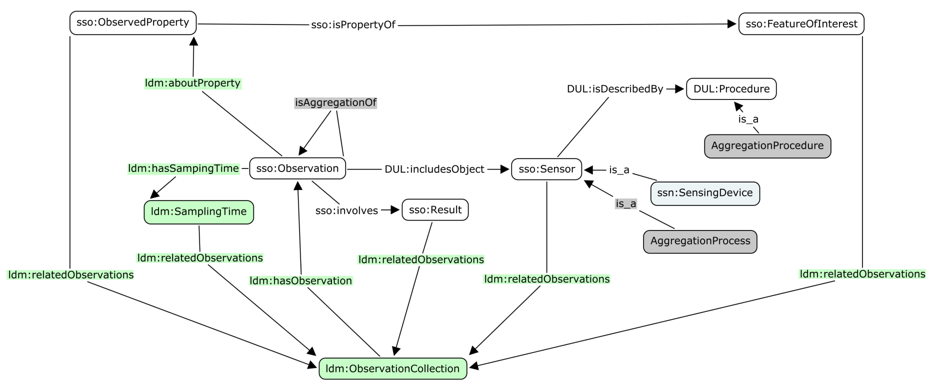

Fig. 1. Partial concept map with the classes and relations of the Linked Data model for aggregated observations, extensions highlighted in grey. The prefix *sso* indicates elements taken from the SSO pattern, DUL indicates elements of DOLCE Ultra Light, *ssn* those of the W3C Semantic Sensor Network ontology (in pale blue), and *ldm* elements of the Linked Sensor Data model.

In order to account for aggregated observations as Linked Data, we extend the SSO pattern with the following elements:

– *isAggregateOf*: a relation that allows one observation to be aggregated out of others; e.g., an observation of daily $PM10$ concentration being an aggregate over hourly measures, or an observation of $PM10$ in Münster, Germany being an aggregate over various Point of Interest (POI) measures.
– *SensingDevice*[13]: a sensor, which is a physical measuring device; e.g., a particular air sampler including a special filter $PM10$.
– *AggregationProcess*: a sensor, which implements a concrete aggregation procedure (see below), for example the process that calculates regional $PM10$ concentrations based on several $PM10$ concentration observations and additional calibration parameters.

[13] The concept of a *SensingDevice* is also captured as part of the W3C SSN-XG ontology. However, it is not part of the SSO pattern, which is applied in our work. We decided to introduce the *SensingDevice* in particular opposed to the notion of the *AggregationProcess* in order to stress the difference between a single physical measurement instrument and the aggregation process that combines multiple sensory inputs to a new observation.

– *AggregationProcedure*: the specific procedure used for aggregating several observations into one; e.g., calculating the MEAN of 24 hourly observations of $PM10$ concentration, or a Kriging interpolation method

The relations between the classes presented in Figure 1 act as links in our model and define the multiple navigation paths and external references; see also [3]. The above mentioned extensions allow for the generation of aggregated observations together with an explicit mentioning of the applied aggregation method, such as MIN, MEAN, or MAX calculations over a temporal series. This also allows for linking aggregated observations back to finer grained observations (discussed in Section 3.3). This new model can be used as URI scheme and query filter to enable the Restful Linked Data SOS to serve aggregated data as well.

3.2 Effects on Links from and to Observations

Aggregating linked observations affects the links from and to the observations. Questions like 'Are the links to a feature of interest still valid, if observations taken at specific points are spatially aggregated to an area?' or 'Which new links can be established after aggregation of an observation?' need to be answered. First of all, the links which are defined in our observation ontology need to be checked for consistency and changed, if necessary[14]. Table 1 shows examples of objects (i.e., link targets) of the links from observations before and after aggregation of point observations to an area in space and a period in time.

Independent of the concrete example, for each aggregation the target of the *ldm:hasSamplingTime* link changes from an instant in time (original observations) to a period in time (aggregated observations), if the observations are aggregated temporally. Also, the *DUL:includesObject* link will always point from the aggregated observation to an instance of an *AggregationProcess* instead of pointing to a specific *SensingDevice* from the original observations. In environmental applications, the *ObservedProperty* is usually a continuous phenomenon, which is sampled at certain locations in space or time, e.g., $PM10$ concentration. If only spatial and/or temporal aggregations are applied, the *ldm:aboutProperty* remains the same. In case of a thematic aggregation (see Section 2.1), the *ObservedProperty* changes. An example is the blizzard as a combination of high wind-speed, heavy snowfall, and low temperatures: the *ObservedProperty* of the blizzard observation is the phenomenon of the blizzard, whereas the original observations point to the phenomena wind-speed, snowfall and surface temperature. Similar examples could be constructed for landslides or forest fires.

Though the *ObservedProperty* might be unchanged during an aggregation process, the *sso:isPropertyOf* link changes, if the observations are aggregated in

[14] Here, changing links means that triples of the original observations might be removed and replaced by other triples in the aggregated observations for the same relationship. For example, the hasSamplingTime relationship usually links to a point in time in the original observations, but to a time period in the aggregated observations, if the observations are aggregated temporally.

Link in Ontology	**Object Before Aggregation**	**Object After Aggregation**
ldm:hasSamplingTime	TimeInstant *(08/05/2011; 23:15 CEST)*	TimePeriod *(one day)*
DUL:includesObject	SensingDevice *(air sampler)*	AggregationProcess *(block kriging of PM10 measures)*
ldm:aboutProperty	ObservedProperty *(PM10)*	ObservedProperty *(PM10)*
sso:isPropertyOf	SamplingPoint *(N 51 57.466 E 007.37.433)*	GeospatialRegion *(area of Münster)*
sso:involves	MeasurementValue *(110)*	Aggregate derived from multiple Measurement Values *(70)*

Table 1. Object of links from observations before and after an aggregation of point measurements to an area in space and period in time; examples are given in parenthesis.

space. For example, aggregating the point measurements to an area causes the *FeatureOfInterest* to change from a sampling point to an upper level feature like the area of the city of Münster. Finally, the *sso:involves* link points to an aggregate computed during the aggregation process. Originally, the *sso:involves* link has pointed to the measurement values from the source observations. Besides changing the original links, additional links might be added pointing to the aggregated observation. As introduced in our model, the *isAggregateOf* link points from an aggregated observation to the original observations. Furthermore, other observation collections might contain the aggregated observation resulting in new *ldm:hasObservation* links to the aggregated observation. Also, other higher level features like cities, administrative areas, etc. might be linked to the aggregated observations.

Formalizing the changes of links during aggregation is challenging and often domain specific. However, we consider the identification and formalization of such changes as crucial to provide a (semi-)automated aggregation of observations in the Linked Data cloud in future and are currently working on such a formalization.

3.3 Provenance in Aggregated Linked Observation Data

In a Linked Data context where different communities might have interest in interlinking their datasets, it is important to publish trust-able datasets. Provenance information favors trustworthiness of data because users are able to analyze the historic changes and reproduce them [24]. Especially when *aggregating* observations in Linked Data, it is important to be able to retrieve information about the original observations as well as the aggregation process that has been applied. Figure 2 shows a provenance graph that illustrates how provenance information about the aggregation process and original observations is provided in our Linked Data model and how this can be mapped to the concepts of the OPM and the Provenance Vocabulary. The reason to extend our model instead of re-using an existing solution lies in the fact that most of the provenance information needed for sensors and observations is already available, thus we avoid redundancy.

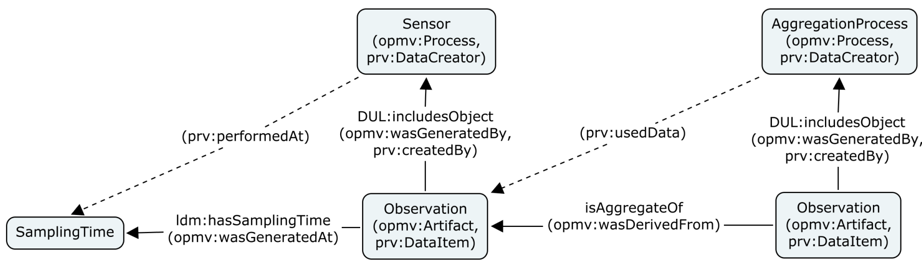

Fig. 2. Provenance graph illustrating how provenance information about the aggregation sensor and original observations is provided in our Linked Data model. The corresponding entries of the OPM vocabulary (with prefix *opmv*) and of the Provenance Vocabulary (with prefix *prv*) are shown in brackets. In case the concepts can not directly mapped but the provenance information can be retrieved following other links in the model, we use dotted lines.

First of all, the *isAggregateOf* allows to trace the aggregated observations back to the original observations. Hence it can be mapped to the *opmv:wasDerivedFrom* relationship in the OPM. Though the *isAggregateOf* relationship cannot be directly mapped to a relationship in the Provenance Vocabulary, it provides basically the information that is provided by the *prv:usedData* link from a *prv:DataCreator* to a *prv:DataItem*. Information about the aggregation process that has created an aggregated observation is provided by the *DUL:includesObject* link from the aggregated observation to the *AggregationProcess*. This link can be mapped to the *opmv:wasGeneratedBy* relationship in the OPM and to the *prv:createdBy* relationship of the Provenance Vocabulary. The *ldm:hasSamplingTime* attribute provides a link to the time at which the value represents a physical phenomenon in the world. In case of observations taken by a physical sensor this corresponds to the time when the observation has been taken. However, if the observations gathered by physical sensors are aggregated by an *AggregationProcess*, the *SamplingTime* is a time period representing the value for which the aggregate is valid. This is no longer the time when the observation has been produced (time of aggregation). Thus, for aggregated observations, an additional time link might be added providing this information. Similarly, additional links might be provided for the *opmv:wasControlledBy* and the *opmv:used* relationships of the OPM, which we did not yet include, as we focus on retracing the observations and not on the users which are aggregating or using the observations.

4 Discussion

The presented research is in line with the theoretical challenges in Sensor Web research, which have been identified during an expert meeting in 2010 [25], addressing the challenges of interoperability and integration of sensor based system and model based systems. Our extension of the SSO design pattern as described in Section 3.1 allows to expose aggregated observations as Linked Sensor Data. This goes beyond the approaches available for providing Linked Sensor Data [14,15,16] which are focusing on providing non-aggregated observations. In our approach, we follow an observation-centric viewpoint that an aggregated observations is still an observation about a quality in the world and thus can be modeled as such. However, further discussion is needed whether the aggregation process can still be modeled as a sensor or has to be distinguished from the concept of sensors.

Our model also allows to retrace the aggregated observations back to the original observations and to retrieve information about the aggregation process applied, thus providing provenance information about the aggregated observations (see Section 3.3). Instead of providing additional metadata as in other approaches described in Section 2.3, we show how the provenance information can be directly retrieved by using the links established in our Linked Data model for (aggregated) observations. For example, Patni et. al. [10] present an approach for provenance in Linked Sensor Data where a separate provenance ontology has been defined. In contrast, we aim to avoid duplication of, for example, information about which sensor has created an observation at what time. This information is already contained in the existing sensor and observation ontologies. We rather show how the relationships of the observation ontologies providing this information can be mapped to relationships of well-established provenance models like the OPM or the Provenance Vocabulary. In order to enable the integration of observations in tools relying on this common provenance models, the observations can either easily be translated to such models or additional triples can be added in the observation set. However, in both approaches, this causes redundant information which might cause problems dealing with large datasets which is common in environmental sciences. Opposed to the general approach for providing provenance information, e.g., about triples in the Linked Data cloud [20], we do not yet consider provenance information about the instances of objects and links according to our Linked Data model, e.g. *Who has created an observation triple in the Linked Sensor Data at which time*. To provide such information, we think that the general approaches for data provenance in the Web can be utilized.

Both, sensor observations and aggregates provide estimations for physical phenomena occurring in the world. As it is not possible to observe all relevant aspects in reality, observations can only represent reality to a certain degree and thus are uncertain about reality. In studies dealing and using observations, it is crucial to account for the uncertainty. This is usually referred to as uncertainty propagation [26]. Aggregation is one mean to adjust the uncertainty in estimations. The more the data is aggregated, the less uncertainty is in the data. At the moment, we do not yet explicitly account for uncertainty in the presented work. Investigations how uncertainty can be propagated in observation processing workflows in the Web are currently ongoing within the European research project UncertWeb[15] [27]. We plan to adopt their approaches and add the uncertainty to our Linked Data model.

While we are providing the model for exposing aggregated observations as Linked Data and we discuss the effects on links from and to observations during aggregation (Section 3.2), we have not yet addressed the technological aspect of executing aggregation processes on Linked Sensor Data. However, we are currently working on extending our Spatio-Temporal Aggregation Service to also deal with Linked Data serialized as RDF. This also leads to the question to what degree observations should be aggregated before exposing them as Linked Data in order to reduce the amount of triples or whether observations can/should be provided at different aggregation levels as Linked Sensor Data. For example, providing high resolution sensor data as Linked Data might lead to a huge amount of triples which might cause performance problems. Thus, it might be better to aggregate the observations before and then expose them as Linked Data.

5 Conclusions and Outlook

In this paper, we identify the need for spatial, temporal, and thematic aggregations of sensor observations and their propagation as Linked Data for an easy integration with other data sources. Aggregates of sensor observations (e.g., the monthly average fine dust concentration in a city) can be much easier utilized in applications. Facilitating the integration of such aggregated observations by providing them as Linked Data enables their utilization among different applications. We achieve this by: (1) extending the SSO ontology design pattern to accommodate aggregation information and including concepts such as *AggregationProcedure* or *AggregationSensor*; (2) describing how links from point observations change after aggregation (e.g., feature of interest may change from a sampling point to a city area); (3) supporting the provenance information in the model through enabling retraceability to original observations and introduce relations such as *isAggregationOf*.

Our future work will follow these lines. Aside from our ongoing implementation work, we plan to exploit the combination of the proposed approach with event detection mechanisms and stream processing. Therefore, we are planning to combine the extension of the SSO ontology pattern presented in this paper with our previous work on *sensor plug & play* [28]. In that work, we designed a framework that enables the on-the-fly integration of sensors and Sensor Web services by determining the semantic matching between sensor characteristics and service requirements. This framework can also be put to use in on-stream processing for the dynamic fusion of incoming data streams of multiple sensors to produce aggregated observations. This is similar to approaches such as [29],

[15] http://www.uncertweb.org

but also allows the creation of new, combined phenomena. A basic example is the combination of *temperature* and *conductivity* data streams measured by underwater sensors to derive a stream of *salinity* observations.

Furthermore, we are planning to extend our approach developed with representations for uncertainty as described in the Uncertainty Markup Language [30]. Our provenance information currently provides information about the aggregation procedure applied, its implementation, and about the original observations that have been used to derive the aggregated observation. In future, it has to be explored how to add additional provenance information about providers and users of the (aggregated) observations.

Our approach of aggregation in Linked Data also allows to utilize the semantics of the links and the objects. First of all, the changes to links as described in Section 3.2 can be translated into rules to check whether adding or removing links is allowed or not. In a next step, the process of adding and removing links during aggregation of observations might be automatized. Furthermore, the semantic reasoning can be used to decide, whether a certain aggregation procedure can be applied to a certain set of observations. Considering, e.g., a set of water level measurements along rivers in Germany, these should not be interpolated to Germany and the semantics can be used to recommend appropriate or disallow inappropriate aggregation processes. However, in order to realize such a system, an ontology of aggregation processes is needed which we consider to be work done in a longer time frame. We hope that our approach as presented in this paper will contribute towards such a semantically-enabled aggregation system.

Acknowledgments

The presented work is developed within the 52° North semantics community (http://52north.org/semantics), and partly funded by the European projects *UncertWeb* (FP7-248488, http://www.uncertweb.org/), *ENVIROFI* (FP7-284898, http://www.envirofi.eu/), *ENVISION* (FP7-249170, http://www.envision-project.eu/), and through the International Research Training Group on Semantic Integration of Geospatial Information funded by the DFG (German Research Foundation), GRK 1498. We are thankful for discussions with members of the Münster Semantic Interoperability Lab (*MUSIL*), the Spatial Data Infrastructures Unit of the Joint Research Centre (JRC) of the European Commission, and colleagues from the W3C Semantic Sensor Network Incubator Group.

References

1. Larssen, S., Sluyter, R., Helmis, C.: Criteria for EUROAIRNET - the EEA air quality monitoring and information network. Technical Report 12, European Environmental Agency (EEA) (1999)
2. Horalek, J., de Smet, P., de Leeuw, F., Conkova, M., Denby, B., Kurfuerst, P.: Methodological improvements on interpolating european air quality maps. etc/acc technical paper 2009/16. Technical report, EEA (2010)

3. Janowicz, K., Bröring, A., Stasch, C., Schade, S., Everding, T., Llaves, A.: A restful proxy and data model for linked sensor data. International Journal of Digital Earth (2011; accepted for publication)
4. Vega Lopez, I.F., Snodgrass, R.T., Moon, B.: Spatiotemporal aggregate computation: A survey. IEEE Trans. on Knowl. and Data Eng. **17**(2) (2005) 271–286
5. Bierkens, M., Finke, P., De Willingen, P.: Upscaling and Downscaling Methods for Environmental Research. Kluwer Academic Publishers (2000)
6. Stasch, C., Autermann, C., Foerster, T., Pebesma, E.: Towards a Spatiotemporal Aggregation Service in the Sensor Web. Poster Presentation. In: The 14th AGILE International Conference on Geographic Information Science. (2011)
7. Jeong, S.H., Fernandes, A., Paton, N.W., Griffiths, T.: A generic algorithmic framework for aggregation of spatio-temporal data. In: SSDBM '04: Proceedings of the 16th International Conference on Scientific and Statistical Database Management, Washington, DC, USA, IEEE Computer Society (2004) 245
8. Pebesma, E.J., de Kwaadsteniet, J.W.: Mapping groundwater quality in the netherlands. Journal of Hydrology **200**(1-4) (1997) 364 – 386
9. Journel, A., Huijbregts, C.: Mining geostatistics. Academic Press (1978)
10. Patni, P., Sahoo, S., Henson, C., Sheth, A.: Provenance aware linked sensor data. In Kärger, P., Olmedilla, D., Passant, A., Polleres, A., eds.: Proceedings of the Second Workshop on Trust and Privacy on the Social and Semantic Web. (2010)
11. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems **5**(3) (2009) 1–22
12. Lassila, O., Swick, R.: Resource description framework (rdf) model and syntax specification. http://www.w3.org/tr/pr-rdf-syntax/. Technical report, W3C (1999)
13. Prud'hommeaux, E., Seaborne, A.: Sparql query language for rdf. w3c recommendation, http://www.w3.org/tr/rdf-sparql-query/. Technical report (2008)
14. Page, K., De Roure, D., Martinez, K., Sadler, J., Kit, O.: Linked sensor data: Restfully serving rdf and gml. In K., T., Ayyagari, A., De Roure, D., eds.: Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09). Volume Vol-522., CEUR (2009) 49–63
15. Phuoc, D.L., Hauswirth, M.: Linked open data in sensor data mashups. In Kerry Taylor, A.A.D.D.R., ed.: Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09). Volume Vol-522., CEUR (2009) 1–16
16. Patni, H., Henson, C., Sheth, A.: Linked sensor data. In: 2010 International Symposium on Collaborative Technologies and Systems, IEEE (2010) 362–370
17. Na, A., Priest, M.: OGC Implementation Specification 06-009r6: OpenGIS Sensor Observation Service (SOS). Open Geospatial Consortium (2007)
18. Janowicz, K., Compton, M.: The stimulus-sensor-observation ontology design pattern and its integration into the semantic sensor network ontology. In Ayyagari, A., Roure, D.D., Taylor, K., eds.: Proceedings of the 3rd International workshop on Semantic Sensor Networks 2010 (SSN10) in conjunction with the 9th International Semantic Web Conference (ISWC 2010). Volume 668., CEUR-WS (2010)
19. Zhao, J., Bizer, C., Gil, Y., Missier, P., Sahoo, S.: Provenance requirements for the next version of rdf. In: W3C Workshop RDF Next Steps. (2010)
20. Hartig, O., Zhao, J.: Publishing and consuming provenance metadata on the web of linked data. In: In: Proc. of 3rd Int. Provenance and Annotation Workshop. (2010)
21. Park, U., Heidemann, J.: Provenance in sensornet republishing. In: Proceedings of the 2nd International Provenance and Annotation Workshop, Salt Lake City, Utah, USA, Springer-Verlag (June 2008) 208–292

22. Liu, Y., Futrelle, J., Myers, J., Rodriguez, A., Kooper, R.: A provenance-aware virtual sensor system using the open provenance model. In: 2010 International Symposium on Collaborative Technologies and Systems (CTS). (2010) 330 –339
23. Lefort, L., Henson, C., Taylor, K., Barnaghi, P., Compton, M., Corcho, O., Castro, R., Graybeal, J., Herzog, A., Janowicz, K., Neuhaus, H., Nikolov, A., Page, K.: Semantic Sensor Network XG Final Report. W3C Incubator Group Report 28 June 2011. Available at: http://www.w3.org/2005/Incubator/ssn/XGR-ssn/. Technical report (2011)
24. Boose, E.R., Ellison, A.M., Osterweil, L.J., Clarke, L.A., Podorozhny, R., Hadley, J.L., Wise, A., Foster, D.R.: Ensuring reliable datasets for environmental models and forecasts. Ecological Informatics **2**(3) (2007) 237 – 247 Meta-information systems and ontologies. A Special Feature from the 5th International Conference on Ecological Informatics ISEI5, Santa Barbara, CA, Dec. 4-7, 2006 - Novel Concepts of Ecological Data Management S.I.
25. Schade, S., Craglia, M.: A future sensor web for the environment in europe. In: Proceedings of the 24th International Conference on Informatics for Environmental Protection - Enviroinfo2010. (2010)
26. Heuvelink, G.: Error Propagation in Environmental Modelling with GIS. Taylor & Francis (1998)
27. Pebesma, E., Cornford, D., Nativi, S., Stasch, C.: The uncertainty enabled model web (uncertweb). In: Environmental Information Systems and Services - Infrastructures and Platforms. (2010)
28. Bröring, A., Maué, P., Janowicz, K., Nüst, D., Malewski, C.: Semantically-enabled sensor plug & play for the sensor web. Sensors **11**(8) (2011) 7568–7605
29. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: TAG: a Tiny AGgregation service for ad-hoc sensor networks. SIGOPS Operating Systems Review **36** (2001) 131–146
30. Williams, M. and Conford, D. and Bastin, L. and Pebesma, E.: OGC 08-122r2: Uncertainty Markup Language (UncertML) (2009)