Linked Stream Data: A Position Paper

Juan F. Sequeda[1], Oscar Corcho[2]

[1] Department of Computer Sciences. University of Texas at Austin
[2]Ontology Engineering Group. Departamento de Inteligencia Artificial. Universidad Politécnica de Madrid, Spain
jsequeda@cs.utexas.edu, ocorcho@fi.upm.es

Abstract. The amount of sensors publishing data on the Web is increasing as a result of the online availability of Sensor Web platforms that provide support for this task. With such increase in sensor data publication, new challenges arise for the identification, discovery and access to this data. Following the set of best practices to publish and link structured data on the web proposed by the Linked Data community, in this paper we introduce the concept of Linked Stream Data, a way in which the Linked Data principles can be applied to stream data and be part of the Web of Linked Data.

1 Introduction

The cost of deploying sensor networks has been falling in the last years, while their capacity has been increasing steadily. As a result, more sensor networks are being deployed in many different environments (roads, forests, agricultural lands, people, homes, etc.), and the information coming from these sensor networks is being used more often for better situation assessment and decision making.

The amount of information being generated by deployed sensor networks is extremely large. For example temperature sensors can emit their readings every 30 min, while heart rate sensors can send their data to a repository every minute. Having this data available not only internally to legacy applications but also available on the web will provide a new source of knowledge for scientists, decision-makers and other types of users. We can then talk about the worldwide sensor web [1].

However, the availability of this data on the web poses new challenges related to how this data can be discovered, identified and exploited in a range of applications. In other words, there needs to be a way to identify it and describe it consistently and to access it easily.

We believe that there is a good opportunity to apply to sensor data the same principles that have been used for the publication of other types of (more static) data on the (Semantic) Web, in the context of the Linked Data initiative. Basically, Linked Data refers to a set of best practices to be followed in order to publish and link data on the Web, using the following basic principles:

- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.

- When someone looks up a URI, provide useful information, using appropriate standards (RDF, SPARQL)
- Include links to other URIs, so that more things can be discovered.

In the application of these principles, there are several features that clearly differentiate the type of data that is normally published in the Linked Data world from data that is originated from sensor networks. First, sensor-based data is normally related to physical measurements and observations, hence predominantly numerical. Second, time and space considerations have to be addressed: sensors and sensor networks are located in specific geographical positions and these positions are usually important for decision making and for further processing, and the measurements provided are commonly tagged with timestamps (what allows for fragments of sensor-based data to be identified by time windows). Third, data accuracy and uncertainty is another factor to be taken into account. Although these are characteristics of most types of sensor-based data, this is not exclusive from this type of data and could also be the case for other types of data.

In summary, the objective of this paper is to discuss how to apply linked data principles to sensor-based data, in what we call "Linked Stream Data" (LSD). We propose a URI-based mechanism to identify and access Stream Data coming from sensor networks, detailing the main requirements for their creation and the main reasons for taking decisions on their design. Finally, we discuss the open challenges for future research.

2. A Selection of Use Cases

In this section we present a series of use cases where information coming from sensors is used and which may benefit from the availability of this information as linked data.

Linear Road Benchmark [2]. It is a well established benchmark for Data Stream Management Systems. This benchmark specifies a variable tolling system by determining changing factors of car congestion on a highway. Each car on the highway is equipped with a sensor that emits the vehicle's exact location and speed every 30 seconds. The data emitted by the sensors is sent as streams to a central system where statistics are generated about traffic conditions on the highways. This tolling system is designed to discourage drivers to use already congested roads because they have an increased toll. Consequently, it would encourage drivers to use less congested roads because they would have decreased tolls. Although this use case was created in order to test and compare different characteristics of existing data stream management systems, the domain in which it is applied is one that may clearly benefit from its availability as Linked Data.

Heart Sensors. Patients with heart problems can have sensors that monitor their heart rate and current location. The data emitted by the heart sensor can be sent as stream data to the patient's hospital where it is monitored. The hospital can detect if the patient suffers from any heart abnormality in real-time. Furthermore, if a patient is having a heart attack, the hospital can immediately send an ambulance to the patient's exact location. Using common Web protocols for publishing and accessing this data, while preserving security and privacy, can ease the development of such type of emergency management applications.

Environmental Sensors. Environmental researchers need to track a large number of aspects of specific regions. For example, one specific application may be monitoring the temperature and humidity of a region. Usually static sensors are used in this type of applications, because they are monitoring fixed areas, therefore space is not an issue in these types of sensors. However, the accuracy of the measurements made may be relevant.

3. Requirements

In this section we present some of the requirements that can be extracted from the previous use cases and that we consider that need to be satisfied by our proposal for linked stream data.

3.1 Identification and Querying

Resources on the Web are commonly identified by means of URIs. As mentioned in the Linked Data principles, URIs act as unique names (identifiers) for such resources on the web, but can also be extended to objects in the real world. As a consequence, we can propose the use of URIs to identify sensors that are deployed in the real world. Furthermore, URIs may also be used to identify data that is emitted by sensors.

The Linked Data principles also stipulate that these should be HTTP URIs and that once de-referenced, useful information should be provided. Therefore, by getting data back once a URI is de-referenced, the URI acts like a query interface or a RESTful service. Hence, the data returned from a sensor URI should be the metadata about the sensor and from the stream data URI should be the observations of the sensor.

From this discussion we can derive the following set of requirements that will inform our decisions for the creation of Linked Data Streams:

- Req 1: Sensors should be identified by URIs.
- Req 2: Stream Data emitted by sensors should be identified by URIs.
- Req 3: The information returned by a sensor URI should be its metadata.
- Req 4: The information returned by a stream data URI should be the observations of the sensor.

3.2 Time Dimension

Data streams provided by a specific sensor or group of sensors can be identified by a specific moment in time or by a time window. For example, consider a sensor that emits the heart rate of a person. One could identify the exact heart rate of a person at a specific time. Furthermore, one could also identify the series of heart rates emitted by that sensor in a specific window of time.

From this discussion we can derive the following set of requirements:

- Req 5: Stream data should be identifiable at specific moments in time.
- Req 6: Stream data should be identifiable in specific time windows.

- Req 7: Time used to describe time points or time windows should be expressed in a given unit of time (milliseconds, seconds, minutes, etc).

3.3 Space Dimension

Data Streams can also be identified given their spatial context. For example, consider a mobile sensor that emits the heart rate of a person. One could identify the exact heart rate of a person at a specific location. In the case of the linear road benchmark, we may be interested only in data coming from vehicles in a specific segment. Similar to the time dimension case, identifying space in sensors can be done by a specific location (or coordinate in this case), or by a bounding area (similar to the time windows but for space). A bounding area, given a center point, can be a radio, square or polygon.

- Req 8: Stream data should be identifiable at a specific location.
- Req 9: Stream data should be identifiable in a bounding area.
- Req 10: Bounding areas can be defined by a radio, square or polygon.

3.4 Combined Time and Space Dimensions

The notions of time and space are particularly interesting in the case of mobile sensors. In the case of the heart rate sensor, one could identify the exact heart rate of a person at a given time and location. In this case, we consider that the requirements for Time and Space should both be satisfied when identifying stream data from mobile sensors.

- Req 11: Stream data should be identifiable at a specific moment in time and specific location
- Req 12: Stream data should be identifiable in a time window and at a specific location.
- Req 13: Stream data should be identifiable at a specific moment in time and in a bounding area.
- Req 14: Stream data should be identifiable in a time window and in a bounding area.

4. A Proposal for Linked Stream Data

In the previous section, we have presented several requirements in order to identify sensors and stream data coming from sensors on the web. Per our motivation, we believe Stream Data can become part of the Web of Data by defining a URI-based mechanism, following the linked data principles, to identify sensors and stream data, optionally given their spatial and temporal context, which can be provided in many different forms. In this section, we present a proposal of human-friendly URIs[1] to identify sensors and stream data.

[1] It is important to note that the fact of using a human-friendly URI that encodes implicitly its semantics in the URI itself does not mean that the explicit semantics are necessarily provided for them. However, in our proposal we aim at combining human readability and the

4.1 URI for Sensors

Sensors are real objects that can be identified by a URI. The data that is returned after de-referencing the URI is the sensor or sensor network metadata (type of sensor, type of measurements made, etc.).

For example, for the sensor that streams the Heart Rate for a Patient 1, we could use the following URI:

```
http://www.linkeddatastreams.org/sensor/heartrate/1
```

Following the guidelines on "cool URIs" for real-world objects[2], these URIs have to be de-referenceable on the Web and be unambiguous, which is the case in our proposal.

For example, de-referencing the URI that identifies Heart Rate Sensor for Person 1 would lead us to the following RDF document[3]:

```
  @prefix owl: <http://www.w3.org/2002/07/owl#>
  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-
ns#>
  @prefix hr:
<http://www.linkeddatastreams.org/sensors/ontology/HeartR
ateMonitor.owl#>
  @prefix hrsensor:
<http://www.linkeddatastreams.org/sensor/heartrate/>
  @prefix sensor:
<http://www.csiro.au/Ontologies/2009/SensorOntology.owl#>

  hrsensor:1      rdf:type sensor:Sensor ;
                  sensor:measures _measurement .
  _measurement    rdf:type hr:HeartRateMonitor .
```

4.2 URIs for Time

Sensors normally emit data at a certain frequency. In the Heart Rate case, data is streamed every couple seconds. In the Linear Road Benchmark, data is streamed every 30 seconds. Hence it is advisable to have a URI scheme that identifies the observations that the sensor emits at a given time, such as the following:

```
http://www.linkeddatastreams.org/sensor/name/%time%
```

---

provision of explicit semantics when the URI is dereferenced. This last one is the only one that will allow automation of tasks based on semantics. The first one is only for human consumption and should be always considered like that by any application that uses them.

[2] http://www.w3.org/TR/cooluris

[3] Please note that the type of metadata that a sensor can provide is not standardised, and this is only a simple example of how the URI would be dereferenced. The W3C incubator group on Semantic Sensor Networks is currently working on different types of sensor ontologies, which include sensor types, observations and measures, etc.

Furthermore, in many situations it would be also useful to represent time windows (intervals of time), as it happens with most data stream and sensor network query languages (e.g., CQL [3]). Therefore we propose the following URI scheme:

```
  http://www.linkeddatastreams.org/sensor/name/%start
time%,%end time%
```

For example, the following URI identifies the observations from the Heart Rate sensor of Person 1 on July 15, 2009 at 17:00:

```
  http://www.linkeddatastreams.org/sensor/heartrate/1/200
9-07-15 17:00:00
```

De-referencing this URI may lead us to an RDF document like the following:

```
  hrsensor:1   sensor:measures
  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:00>.

  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:00>
     rdf:type hr:HeartRateMonitor;
     hr:heartRate "74";
     hr:timestamp "2009-07-15 17:00:00"^^xsd:dateTime .
```

And the following URI identifies the observations from the Heart Rate sensor of Person 1 between July 15, 2009 at 17:00 and July 15, 2009 at 18:00:

```
  http://www.linkeddatastreams.org/sensor/heartrate/1/200
9-07-15 17:00:00, 2009-07-15 18:00:00
```

For example, de-referencing this URI may lead us to the following RDF document[4]:

```
  hrsensor:1  sensor:measures
  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:00>.
  hrsensor:1  sensor:measures
  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:05>.
  hrsensor:1  sensor:measures
  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:10>.
```

Note that if %start time% and %end time% are the same, then the time window is 0, hence the URI would be the same as the URI with one given time. For example:

[4] The three measures hereby shown as an example may also come bundled in an RDF list.

```
  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:00, 2009-07-15 17:00:00>
  owl:sameAs
  <http://www.linkeddatastreams.org/sensor/heartrate/1/20
09-07-15 17:00:00>
```

4.3 URIs for Space

A sensor may also emit its current location together with the rest of data, if it is mobile, or may have as part of its metadata information about the location where it is place, in the case of static ones. In the Heart Rate and the Linear Road cases, sensors are mobile, hence their coordinates depend on the sensor location.

The Linked GeoData project has already proposed a way of representing spatial dimensions as Linked Data. Therefore we follow their guidelines. A URI to identify spatial dimension from a sensor would be:

```
  http://www.linkeddatastreams.org/sensor/name/%latitude%
,%longitude%/%radius%
```

For example, the following URI identifies the car position emitted from Car 1 in a 1 meter radius from the coordinates 50.60242, -2.5225.

```
  http://www.linkeddatastreams.org/sensor/car/1/50.60242,
-2.5225/1
```

For example, de-referencing this URI would lead us to the following RDF document:

```
  @prefix carsensor:
<http://www.linkeddatastreams.org/sensor/car/>
  carsensor:1  sensor:measures
  <http://www.linkeddatastreams.org/sensor/car/1/50.7,   -
2.6>.
  carsensor:1  sensor:measures
  <http://www.linkeddatastreams.org/sensor/car/1/51.2,   -
3.3>.
  ...
```

This leads us to have URIs that identify observations from sensors are specific locations. If this type of URI would be de-referenced, it would lead us to the following RDF document:

```
  carsensor:1  sensor:measures
  <http://www.linkeddatastreams.org/sensor/car/1/50.7,   -
2.6>.

  <http://www.linkeddatastreams.org/sensor/car/1/50.7,   -
2.6>
```

```
rdf:type    lr:CarLoc;
lr:speed "60";
 hr:timestamp "2009-07-15 17:00:00"^^xsd:dateTime .
```

A same location can emit different observations, and this would be differentiated through time, because we assumed that two or more observations of a sensor cannot be emitted at the same time.

In these examples, we have used a radius as the bounding area. Other possibilities can be the use of a bounding box or a bounding polygon.

4.4 URIs for Time and Space

Now that we have proposed how to represent Time and Space in URIs separately, we propose a way to represent them together. As discussed before, this type of URI makes more sense for sensors that are mobile. Given that we have proposed two ways of representing time in URIs (specific time and time window), we therefore propose the following two schemes for representing time and space in URIs (please note that temporal attributes go before the spatial attributes, however, there is no strong reason for this and it could be done in a different order):

```
http://www.linkeddatastreams.org/sensor/name/%time%/%latitude%,%longitude%/%radius%
http://www.linkeddatastreams.org/sensor/name/%start time%,%end time%/%latitude%,%longitude%/%radius%
```

The following URI identifies the observations of the Heart Rate sensor of Person 1 on July 15, 2009 at 17:00 when it was in a 1 meter radius from the coordinates 50.60242, -2.5225:

```
http://www.linkeddatastreams.org/sensor/heartrate/1/2009-07-15 17:00:00/50.60242,-2.5225/1
```

The following URI takes in account a time window and identifies the observations of the Heart Rate sensor of Person 1 between July 15, 2009 at 17:00 and July 15, 2009 at 18:00 when it was in a 1 meter radius from the coordinates 50.60242, -2.5225:

```
http://www.linkeddatastreams.org/sensor/heartrate/1/2009-07-15 17:00:00,2009-07-15 17:00:00/50.60242,-2.5225/1
```

5. Related Work

To our knowledge, there has not been an approach that applies the linked data principles to stream data. However, several other approaches take in account some of the issues that we have presented in this paper.

First, we have presented a set of human-friendly URIs that can be used to identify sensors and sensor data. However, these human-friendly URIs may be written in many other different ways. For example, Davis [4,5] proposes a different way of representing intervals in URIs. Instead of having the start and end time in the URI, one could write the start time with the duration. This approach is feasible and we consider that a user should decide what the best approach is.

The approach presented by Pfeiffer et al [6] specifies a syntax for addressing time intervals within time-based Web resources through URI queries and fragments. However, these types of URIs do not follow the Cool URIs approach used in Linked Data.

Hausenblas et al [7] presented an approach for applying the Linked Data principles to multimedia fragments. This approach focuses on addressing multimedia fragments through a URI-based mechanism. The main motivation behind it is the ability to send only the multimedia fragment that is wanted by the user, instead of sending the whole multimedia file and then having the user multimedia client find the relevant fragment, which is what it is currently done. Furthermore, this approach also describes metadata about multimedia fragments, which enables it to be part of the Linked Data cloud.

6. Conclusions, Open Challenges and Future Work

In this position paper we have described our proposal for the generation of Linked-Data-observant URIs for sensors and sensor data, so that this wealth of information could be easily included in the Linked Data cloud. We have covered the two most relevant issues that have to be dealt with when considering this type of data, time and space, and we have tried to be compliant with existing proposals that have addressed the use of time and space in URIs, including Linked Data efforts in the of geolocalition-related information.

One important aspect to be considered is that by just generating human-readable URIs that encode internally time and space constraints we are not providing the explicit semantics of those URIs so that they could be adequately consumed by semantic-aware systems. This means that when dereferencing these URIs, the results obtained (e.g., in RDF) should also reflect that information as part of the information provided.

Another important consideration that has to be made is that this way of defining URIs allows for the easy creation of REST-like query interfaces to sensor data. The time and space constraints expressed in the URIs can be transformed into stream query languages that will allow performing transformations into RDF on the fly. However, architectural decisions about how to make these transformations still need to be made, and corresponding implementations will need to be done as well.

A final aspect that has not been considered in our proposal, and that is still open, is the handling of uncertainty in the data that is provided when dereferencing URIs or when expressing them. Furthermore, the amount of Linked Data that can be produced will result into scalability issues that need to be addressed. We have discussed that this is an important feature of this type of data and should be handled somehow.

Finally, it is important to note that other sources of information, such as RSS feeds, social network feeds (Twitter, Facebook, etc.) may be also considered as stream data, even if used for different types of applications. Even though in this paper we are considering stream data emitted from physical sensors, our proposal may be applied to these other types of stream data.

Acknowledgements

This work is supported by the EC project SemsorGrid4Env (http://www.semsorgrid4env.eu/). We appreciate the discussions held with the project members about the different requirements and design decisions taken, which have allowed us to improve our proposal.

References

1. Magdalena Balazinska, Amol Deshpande, Michael J. Franklin, Phillip B. Gibbons, Jim Gray, Mark Hansen, Michael Liebhold, Suman Nath, Alexander Szalay, Vincent Tao, "Data Management in the Worldwide Sensor Web," IEEE Pervasive Computing, vol. 6, no. 2, pp. 30-40, Apr.-June 2007
2. Arasu, A., Cherniack, M., Galvez, E., Maier, D., Maskey, A., Ryvkina, E., Stonebracker, M., Tibbetts, R. Linear Road: A Stream Data Management Benchmark. In Proceedings of Very Large Databases Conference 2004.
3. Arasu, A., Babu, S., Widom, J., The CQL continuous query language: semantic foundations and query execution. VLDB Journal 15(2). 2006
4. Davis, I. http://placetime.com/instant/gregorian/
5. Davis, I. http://placetime.com/interval/gregorian/
6. Pfeiffer, S. Parker, C. Pang, A. (2005) Specifying time intervals in URI queries and fragments of time-based Web resources. Internet Draft. http://www.annodex.net/TR/draft-pfeiffer-temporal-fragments-03.html
7. Hausenblas, M. Troncy, R. Buerger, T. Raimond, Y. (2009) Interlinking Multimedia: How to Apply Linked Data Principles to Multimedia Fragments. Linked Data on the Web Workshop at WWW 2009.