From Sensor Data Streams to Linked Streaming Data: a survey of main approaches

Kathrin Rodriguez Llanes[1], Marco Antonio Casanova[1], Noel Moreno Lemus[2]

[1] Pontificia Universidade Católica do Rio de Janeiro, Brazil
kllanes@inf.puc-rio.br, casanova@inf.puc-rio.br
[2] Laboratório Nacional de Computação Científica, Brazil
noelml@lncc.br

Abstract. Today, large amounts of data are produced by sensor networks. They are continuously producing data streams about real world phenomena. However, these data streams are generated in raw and different formats, lacking the semantics to describe their meanings, which imposes barriers to accessing and using them. To tackle this problem, several solutions using Linked Data Principles have been proposed. In this article, we survey the main solutions developed by the research communities for publishing data streams in the Web of Data. The major contributions of the article are the identification of the strengths and limitations of these solutions and, over that basis, the main steps that someone should follow to publish data streams in a manner that anyone can use them, with a minimal understanding of the details. We also highlight the main challenges that emerge from this survey, concluding with a list of research tasks for future work.

Categories and Subject Descriptors: H.2.5 [**Database Management**]: Heterogeneous Databases; C.2.3 [**Computer-Communication Networks**]: Network Operations

Keywords: data streams, linked data, semantic web, sensor data publishing

1. INTRODUCTION

In recent years, data sensor networks have been deployed in various domains (medical sciences for patient care using biometric sensors, wildfire detection, meteorology for weather forecasting, satellite imagery for earth and space observation, agricultural land, etc.). The sensors are distributed across the globe, capturing and continuously producing an enormous amount of data streams about a number of real world phenomena.

However, typically, the data streams produced by sensor networks are in raw and different formats, lacking the semantics to describe their meaning. This failure intensifies the current traditional problem "too much data and not enough knowledge" [Sheth et al. 2008] and imposes barriers to accessing and using sensor data in applications and linking them with other related data sources.

To tackle this problem, several solutions using Linked Data Principles [Berners-lee 2006] have been proposed. They allow to integrate sensor technologies with Semantic Web technologies in order to publish sensor data streams in an enriched and standardized way, so that they can be accessed and consumed by external applications. The publication process consists of transforming the data streams into linked streaming data following the Linked Data Principles.

During the process of publishing data streams in the Linked Open Data (LOD) cloud, the time component plays a key role, which substantially changes the way of data processing compared with the publishing of static data.

The publishing of static data in the LOD cloud is composed of several activities: specification, modelling, generation, publication and exploitation. The specification refers to a preliminary set of tasks to identify and analyze the data to be published. Then, we need to select the ontology or ontologies to be used for modelling and semantically describing data. After that, the data are transformed to standard representation in RDF[1] format and linked to external data sources through the generation activity. This activity ends with a meaningful and enriched tripleset. During the publication activity, this enriched tripleset is stored and published in a triple store to be consumed later. Once the data are published on the Web of Data, they would be queried and consumed. The activity that takes care of these tasks is called exploitation, which is the main purpose of the publication process as whole.

In multiple domains, the time component is critical to make the right decisions quickly. In particular, it implies that data stream processing must be done in real-time. This means that data from sensor observations should be processed on-the-fly with a minimum delay. To fulfill this requirement, significant modifications to the traditional static data publishing process should be made, such as the incorporation of data compression, data stream abstraction, continuous queries and the generation of links in real time, among others.

Several efforts have been developed for publishing data streams on the Web of Data. Some take into account the real-time and others do not consider it. In this paper, we survey the main works available in the research literature, identifying their weaknesses and strengths. Based on an analysis of the strengths and weaknesses, we propose a set of research tasks to be faced in the near future.

The remainder of the article is organized as follows. Section 2 exposes a list of motivational scenarios demonstrating the need to make sensor data available, accessible and queryable in real-time. Section 3 describes in detail the main steps that someone should follow to publish data streams in a way that anyone can use them with minimal understanding of the data details. Section 4 reviews the most relevant approaches proposed to publish data streams on the Semantic Web following the Linked Data Principles. Section 5 discusses the lessons learned and lists open challenges that emerged from this survey. Section 6 concludes the paper and presents our future research directions.

2. MOTIVATING SCENARIOS

There are dissimilar scenarios in which researchers have been inspired to integrate physical data produced by sensors and virtual data from social networks and publish them in the LOD Cloud in real-time. We will describe some of them below.

To know what weather station is detecting a hurricane or other atmospheric phenomena affecting the population. In the last couple of decades, atmospheric phenomena affecting the planet have been increasing in number and the damage caused. For this reason a large number of weather organizations collect a vast amount of data from sensors deployed around the world in order to predict future catastrophic situations and thus take measures to prevent further disasters.

To determine which route a driver should follow to arrive to a place in the shortest possible time. The driver drives following the routes suggested by his particular GPS car navigation system. The route suggestions will be based on composite set of data accessed from LOD's datasets such as information about the hilly surrounding area (from Geography LOD datasets), information about nearby road works (from Government LOD datasets), and information about ongoing social events in the locale (from Media LOD datasets) [Hübner et al. 2015]. For example, consider the case of city where is happening a large event that requires city-wide mobilization of urban resources: Olympic Games to be held in Rio de Janeiro, Brazil in 2016, and the large crowd arriving in one area in Rio de Janeiro to watch the last match of football that defines the Olympics winner team leading to traffic

[1] www.w3.org/RDF

chaos. A driver that is going from a given neighborhood to the Maracanã Stadium wants to know which route he should follow to arrive to the stadium in the shortest possible time. By integrating and correlating partial observations from multiple data sources, physical sensors and virtual sensors (Twitter, Facebook, etc), a system could infer the appropriate road and suggest it to driver.

To determine when to walk to the next bus stop to arrive in time to go some place. Some users want to go to an event (identified via e.g., Eventful) in the metro area of a city. Based on the users current location, they want to know when to be at the next bus stop to arrive in time, for the event. The scenario includes a real-time visualization of the area that is augmented with current data from multiple sources to help a user make well-informed and timely judgments. To solve this problem we need to integrate new sources rapidly [Zhang et al. 2015].

Smart Cities. Smart Cities represent a specific field where the integration of Sensor Web technologies and Semantic Web technologies can find a concrete application. The final goal of the research in this area is to process and understand the information relevant for the life of a city and use it to make the city run better, faster, cheaper and sustainably [Cheng et al. 2015].

Remote Health Monitoring. Remote health monitoring allows remote patient monitoring, focusing on collecting information from multiple sources (e.g., sensors for monitoring the heart rate or blood pressure). Also by applying a reasoning engine, it is possible to understand the context or situation of the patient and to guide the decision making process. In this context, one of the main challenges is the integration of data from multiple sources, e.g., it may be useful to know the current activity of the patient to understand if the measured heart rate is too high [Mukhopadhyay 2015].

Integrated Presence management. The aim of integrated presence management is to combine sources of virtual presence, for example, calendar information, online status in Skype and chats, in IP telephone systems, collaborative environments, etc., with physical presence information from different relevant sensors data (e.g. Geographic Position Sensor, Wireless LAN, Bluetooth, RFID, activity detection sensors, and noise sensors etc) from the LOD cloud for obtaining an integrated view of an entity's presence and availability. This data can be used both in simple and more complex scenarios. For example, if data from the user's calendar is combined with his physical location, then appointments can be automatically changed if the system infers that all participants are not able to get to the meeting location on time. On the other hand, depending on the physical location or activity of a particular user detected by several sensors (e.g. audio sensors or RFID readers in the user's office etc), his online status on Skype, on the phone and in a chat client can be changed automatically. In more complex scenarios the integrated presence management can be used too, if sensor data is flexible and transparently integrated with other sources. Linked Data can be used to efficiently establish links between sensors data and LOD datasets and to make them globally accessible.

3. SENSOR DATA PUBLISHING ON THE SEMANTIC WEB

Sensor networks employ various types of hardware and software components to observe and measure physical phenomena and make the obtained data available through different networking services. Applications and users are typically interested in querying various events and requesting measurement and observation data from the physical world. Through the process of publishing sensor data on the Semantic Web, knowledge is added over raw sensor data in order to satisfy the high-level information requested by the queries.

The process of publishing sensor data in the Semantic Web encompasses three main stages: mapping and conversion from data streams to RDF streams, storing RDF streams and linking them with related data sources existing in the LOD cloud. To carry out this process a set of important tasks are required, such as: *(i)selection of ontologies to semantically describe data streams*; *(ii) defining the mapping language to do the conversion*; *(iii) selection of continuous query languages*; *(iv) choosing the appropriated datasets from the LOD cloud and creating data linkages.* See Figure 1. To support

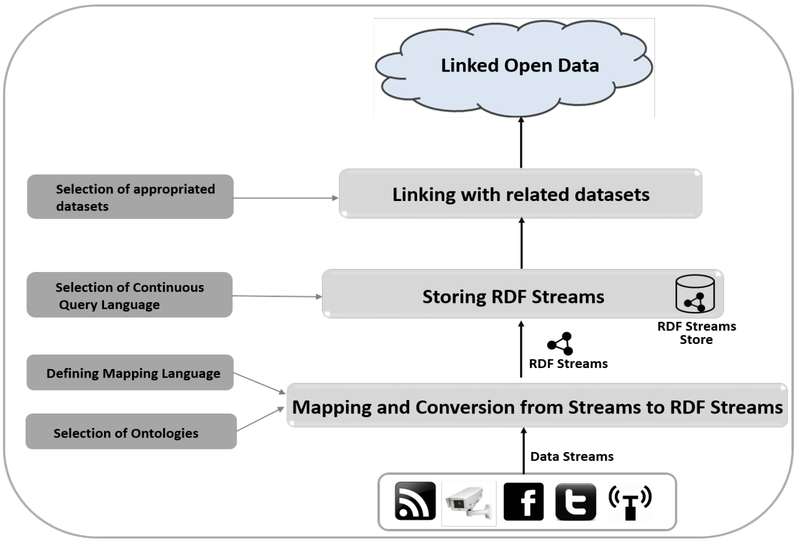

Fig. 1. Data Stream Publication Process

the complete process a stream publishing framework is being developed.[2]

3.1 Selection of ontologies

With the development of semantic sensor networks, a number of ontologies describing the sensor network domain have been brought forth in the past years. A detailed survey was performed by Compton et al. [2009], where eleven sensor network ontologies were analyzed. Therefore, considering the need of standardization regarding sensor networks ontologies, the Semantic Sensor Network Incubator Group from W3C was formed, with the purpose of developing ontologies for sensor networks and searching for appropriate methods for enhancing available standards with semantic technologies. As a result of the efforts of this group, the Semantic Sensor Network (SSN) ontology [Compton et al. 2012] was defined, and it can describe capabilities, measurements, and resultant observations from sensors.

The W3C Semantic Sensor Network Incubator Group also developed a methodology to perform semantic annotations over the data generated by sensors following the standards defined by the Open Geospatial Consortium (OGC). These standards help describe observed phenomena such as space, time, and theme.

Spatial metadata provide information regarding the sensor location and data, in terms of either a geographic reference system, local reference, or named location. Temporal metadata provide information regarding the time instant or interval when the sensor data is captured. Thematic metadata describe a real world state from sensor observations, such as objects or events. All these metadata play an essential role in managing sensor data and provide more meaningful descriptions and enhanced access to sensor data. Both projects, the SSN ontology (SSNO) and the proposed methodology, developed by W3C Semantic Sensor Network Incubator Group, facilitate the enrichment and the semantic fusion of the stream data applications, since they allow to publish the streaming data and also integrate them with other related datasets. Sometimes, sensor ontologies are not able to provide all the semantics needed by a scientific system, and thus additional ontologies are often required.

[2]https://github.com/nmlemus/streams2LSD

3.2 Defining the mapping language

Several languages have been proposed by the Semantic Web research communities for expressing customized mappings from relational databases to RDF datasets. Such mappings provide the ability to view existing relational data in an RDF data model and they are expressed in a structure and target vocabulary of the author's choice mapping. D2R [Bizer 2003], R2O [Barrasa et al. 2004] and R2RML [Consortium 2012] represent some of the standard languages. They are useful in transforming static relational data to RDF, but present some disadvantages when facing the challenge of converting data streams to RDF streams.

Despite the existence of this gap, solutions for streaming data mapping and querying using ontology-based approaches have been little explored. Calbimonte et al. [2010] presented an extension of R2O called S2O for the data stream to RDF mapping. Also, Harth et al. [2013] developed an extension for R2RML with the same purpose.

Zhang et al. [2015] designed a XML-based mapping language (named SASML) used to annotate sensors and sources. They also proposed an algorithm (named SDRM) to transform sensor data to RDF conforming to SSN ontology, based on the predefined SASML mapping file.

The most mature mapping languages are S2O and the R2RML extension, and their main advantage is that they allow using multiple ontologies for the semantic annotation process. Meanwhile, SASML is easier to understand, but is designed specially for the mapping between sensor data and SSN ontology(the standard for semantic sensor network).

3.3 Selection of continuous queries languages

Languages such as SPARQL are designed to execute queries over RDF triples, but they do not have the functionalities to query RDF streams. To face this challenge, continuous RDF query languages have been proposed.

Barbieri et al. [2009] introduced the Continuous SPARQL (C-SPARQL) as the extension of SPARQL for querying RDF streams. It supports continuous queries, which are registered and continuously executed over RDF data streams, while considering windows of such streams. C-SPARQL is currently not designed to handle large volumes of data and that constitutes their main weakness.

SPARQLstream [Calbimonte et al. 2011] is an extension to SPARQL for RDF streams. It was inspired by the previous proposals: C-SPARQL and SNEEQL [Brenninkmeijer and Galpin 2008], but with some improvements: it only supports windows defined in time; the result of a window operation is a window of triples, not a stream, over which traditional operators can be applied. It uses S2O and R2RML for the definition of stream-to-ontology mappings. Its main disadvantage is that currently it does not support querying on both stream and RDF datasets.

Anicic and Fodor [2011] developed the Event Processing SPARQL (EP-SPARQL). It is a continuous query language that uses a black box approach backed by a logic engine. It translates queries into logic programs, which are then executed by a Prolog engine. EP-SPARQL provides a unified execution mechanism for event processing and stream reasoning, which is grounded on logic programming. The main deficiency of EP-SPARQL is that its performance drops significantly for complex queries.

Phuoc [2013] presented the Continuous Query Execution over Linked Stream (CQELS), an adaptive execution framework for Linked Stream Data and Linked Data. CQELS provides a flexible architecture for implementing efficient continuous query processing engines over Linked Data Stream and Linked Data.

To the best of our knowledge, the most complete approach to continuous queries over RDF streams is CQELS. Despite its scalability issues with respect to multiple concurrent queries, the CQELS engine can achieve better performance compared to other black box systems in relation to order of magnitude. It represents a solution for RDF stream processing built on top of the notion of linked stream data. The solution offers a native way to interpret and implement common stream processing futures (time window operator, relational database such as join and union operators, and stream generation operator) in RDF data stream processing environment.

3.4 Choosing related datasets in LOD cloud and creating data linkages

Another important task in the process of data stream publishing on the Semantic Web is the selection of the most suitable triplesets with which RDF streams may be interlinked. It will allow users to take advantage of existing knowledge. However, it is not a trivial task. Usually, users interlink their triplesets mostly with data hubs, such as DBpedia and Freebase, ignoring the most specific and more promising triplesets. To alleviate this problem, some tripleset recommendation tools have been implemented.

Lopes et al. [2013] presented a tripleset recommendation approach using strategies borrowed from social networks. To generate the ranked list, the procedure uses a recommendation function adapted from link prediction measures used in social networks. The tool obtains high levels of recall and reduces up to 90% the number of triplesets to be further inspected for establishing appropriate links.

Caraballo et al. [2014] presented a Web-based application, called TRTML, that explores metadata available in Linked Data catalogs to provide data publishers with recommendations of related triplesets. TRTML combines supervised learning algorithms and link prediction measures to provide recommendations. Its high precision and recall results demonstrate the usefulness of TRTML.

Lopes et al. [2014] developed RecLAK. It is a Web application developed for the LAK Challenge 2014 focused on the analysis of the LAK dataset metadata and provides recommendations of potential dataset candidates to be interlinked with LAK dataset. RecLAK follows an approach to generate recommendations based on Bayesian classifiers and Social Networks Analysis measures. Furthermore, RecLAK generates graph visualizations that explore the LAK dataset over other datasets in the LOD cloud.

The main disadvantage of these tripleset recommendation tools is that they are not able to do the recommendation process on-the-fly, since they are not designed to act in real-time, which represents a gap in the process of sensor data publishing. For this reason, the current solution is to choose the LOD datasets related to each new sensor that will be incorporated into the sensor network, using the tools described above, before sensors start to capture observations.

Once the most suitable related resources are identified within the LOD cloud, the next step is to generate linkages among sensor network data and the related resources, thus completing the process of publishing. To carry out the interlinking task effectively, some algorithms and tools have been developed.

?] proposed a semantic sensor cloud platform-SERAW, which generates the linkages among sensor network data and related data in the LOD cloud by building some functions.

Cuan et al. [2015] proposed a similarity comparison algorithm of graph, based on heuristic property, to generate linkages among sensor network data and related resources in LOD cloud. They use automatic interlinking approach of Linked Data, and also compare similarity of their property concepts.

Between these two proposals, the latter is most effective than the former because, according to the characteristics of sensor network data, and by learning the mapping ideal based on graph similarity, the algorithm implements an heuristic that starts from the property of sensor data to find similar nodes and generate RDF linkages among sensor network data and related resources in LOD cloud at the end.

4. PLATFORMS FOR SENSOR DATA PUBLISHING ON THE SEMANTIC WEB

Although the main goal of Linked Stream Data is to make available the sensor data in the LOD cloud, in real-time, only a few projects have achieved it. In this section, the most popular efforts of research communities to publish sensor data on the Semantic Web will be analyzed. Some of them do not publish sensor data in real-time, which is their main weakness, but it could be a starting point for future work.

4.1 Non real-time approaches

Phuoc and Hauswirth [2009] presented an approach and an infrastructure that makes sensor data available, following the Linked Open Data principles, and enables the seamless integration of such data into mashups. This project publishes sensor data as Web data sources, which can then be easily integrated with other Linked Data sources and sensor data. Also, it allows users to describe and annotate semantically raw sensor readings and sensors. These descriptions can then be exploited in mashups and in Linked Open Data scenarios and enable the discovery and integration of sensors and sensor data at a large scale. The user-generated mashups of sensor data and Linked Open Data can in turn be published as Linked Open Data sources and used by other users.

Patni et al. [2010] presented a framework to make this sensor data openly accessible by publishing it on the LOD cloud. This is accomplished by converting raw sensor observations into a standard representation in Resource Description Framework (RDF) and linking with other datasets on LOD. With such a framework, organizations can make large amounts of sensor data openly accessible, thus allowing greater opportunity for utilization and analysis. They were the first one to add to the LOD cloud a large dataset of sensor descriptions and measurement, by first representing it in Observations and Measurements (O&M) standard.

Barnaghi et al. [2010] proposed a platform called Sense2Web for publishing sensor data description defined by spatial, temporal, and thematic attributes. The platform offers an interface for publishing linked sensor data without requiring from the users a semantic technological background. The sensor observation and measurement data can also be published following similar principles. However, the publishing of observation and measurement data raises other concerns such as time-dependency, scalability, freshness, and latency.

Moraru et al. [2011] proposed a system for publishing sensor data following the Linked Data Principles and providing integration with the Semantic Web. In their proposal, they focused on a single sensor source and used a relational database for storing sensor data, which represents its main deficiency; because a relational database is not prepared to support the continuous arriving of data streams.

Yu and Liu [2015] designed and implemented a system to publish real-world data into linked geosensor data. In their work, they proposed re-using and matching the W3C Semantic Sensor Network (SSN) ontology and other popular ontologies for heterogeneous data modeling in the water resources application domain. Although with in a system a good integration and interoperability is achieved, its main limitation is that it only focuses on the domain of water resources, which means that it can not be used to integrate data from sensors from other domains.

Rezvan et al. [2015] proposed a Semantic Sensor Network Model to be applied in the Internet of Things. It can be applicable in all types of sensors during different stages of sensor data processing for information abstraction. The model processes and converts the raw sensor data into higher-level abstractions to be understood by human and/or machine using semantic web techniques, sensor web technology and machine learning methods. The model increases the abstraction level of the data collected by the sensors of smart objects, but it does not make available the data resulting of data abstraction in the LOD cloud. For this reason, the proposed model does not satisfy the four principle of the Linked Data paradigm:to link data to other data sources on the Web of Data, which represents the major deficiency of this work.

4.2 Real-time approaches

?] proposed an approach to publish data as Linked Data streams. The approach uses C-SPARQL to register and run continuous queries over streams of RDF and C-SPARQL engine to publish the retrieved data as Linked data streams. To represent RDF in RDF streams, they proposed the use of two named graphs: stream graph (S-graph) and instantaneous graph (I-graph). An RDF stream can be represented using one s-graph and several i-graphs, one for each timestamp. The main limitation of this approach is that it is only a prototype, and does not have a full application that supports it.

Le-Phuoc et al. [2011] proposed a Linked Stream Middleware, which is a platform that facilitates publishing Linked Data Stream and making it available to other applications. It provides the following functionalities: wrappers to access stream data sources and transform the raw data into Linked Stream Data; data annotation and visualization through a web interface and life querying over unified Linked Stream Data and data from the LOD cloud. Besides the processing of real-time data, it is also necessary to store the data generated, either for queries defined over a time period or for archiving purposes. These are the main limitations of this approach: the triple storage cannot efficiently handle high update rates; the materializing sensor reading into triples is also inefficient, especially numeric readings, and it also runs into performance issues with complex queries.

Hasemann et al. [] proposed the Platform-independent Wiselib RDF Provider for embedded Internet of Thing (IoT) devices such as sensor nodes. It enables the devices to act as semantic data providers. They can describe themselves, including their services, sensors, and capabilities, by means of RDF documents. The greatest contribution of this proposal is the introduction of Streaming HDT, a lightweight serialization format for RDF documents that allows transmitting compressed documents with minimal effort for the encoding. The platform also allows publishing and sharing sensor data with low cost, less complexity of sensor data integration, and easier access to the integrated sensor data.

Harth et al. [2013] developed a Web architecture that enables (near) real-time access to data sources in a variety of formats and access modalities. Also, it enables rapid integration of new live sources by modeling them with respect to domain ontology and automatically transforming all the arrived data streams from their format (CSV, TSV, JSON) to RDF and publishing them following the Linked Data Principles. This approach is a very good approximation to solve the problem related to the integration of new sensor devices into the LOD cloud, but it is still an immature project.

Soldatos et al. [2014] presented the design principles for an open source platform to integrate heterogeneous geographically and administratively dispersed sensors and IoT services in a semantically interoperable fashion. The main components of the functional framework intend to follow the utility-driven cloud-based computing model. The platform will foster the use of linked sensor data technologies and SSN ontology. The development of the ideas presented in this article will ensure a powerful framework to handle mobile sensor data and develop IoT applications in areas where semantic interoperability is a major concern, but currently there is not a complete application that implements them.

5. LESSONS LEARNED AND OPEN CHALLENGES

5.1 Lessons Learned

The publishing of data streams on the LOD cloud should take into account the following observations:

(1) **Ontology selection**: There are several ontologies designed to semantically describe sensor data that help us during the annotation process. However, sometimes sensor ontologies are not able to provide all the semantics, and additional ontologies are often required.
(2) **Data stream selection**: Before starting the transformation process from data streams to RDF, it is extremely important to make an abstraction of streams to select the most significant data, and not spend time to process those less relevant streams.
(3) **Data compression**: An efficient and lightweight serialization format for RDF should be used in order to transmit compressed documents with minimal effort for encoding.
(4) **Data Integration**: Integrate information from heterogeneous sources (sensor networks and social networks) in order to support decision making in real-time. Integrating sensor data with data from social networks allows capturing human perception, which implies that better decisions are made.

5.2 Open Challenges

In order to integrate sensor technologies with Semantic Web technologies and publish them as Linked Data Streaming in real-time, some efforts have been made. Nevertheless, some challenges are still being faced:

(1) In order to publish and consume data from sensor data streams in real-time, it is necessary a lighter mapping language, capable of guaranteeing mapping on demand, and efficient conversion from sensor data streams to RDF streams.
(2) The conversion of the data streams to RDF streams must be on-the-fly. This restriction captures the idea that the data must be continuously triplified, albeit with limited delay. To fulfill this requirement, techniques for efficient triplification should be developed.
(3) The interlinking process of RDF streams with data sources of the LOD cloud must be on-the-fly. To address the restriction of minimum delay, interconnection techniques should be based on a strategy of pre-processing or caching data to accelerate the creation of links.
(4) The lack of an efficient RDF storage that supports the real-time stream processing. Although the classical RDF storages are efficient to store RDF that will persist over time, they are not efficient to handle the RDF streams, because the RDF streams need to be stored, accessed, and processed on-the-fly.

6. CONCLUSIONS AND FUTURE WORK

Real-time publishing of sensor data streams based on semantic technologies is indeed not only possible, but useful in many application areas. In this paper, we presented a study that covers the main proposed approaches to publish the sensor data in the LOD cloud from 2009 to the present, identifying its main contributions and limitations. We described the main steps that one should follow to publish data streams in a manner that anyone with a minimal understanding of the data details can use them, and we also suggested the most suitable tool to use at every step. Based on the limitations of the current approaches, we are developing a stream publishing framework to cover the gaps. Also, we discussed the ongoing challenges, and to cope some of them we propose the following directions for future work:

(1) Conclude the implementation of the framework that is being developed.
(2) Develop a NoSQL system as a compelling alternative to distributed and native RDF stores for simple workloads. Considering its strengths, its very large user base, and the fact that there is still ample room for query optimization techniques, we are confident that NoSQL databases will present an ever growing opportunity to store and manage RDF data in the LOD cloud.

REFERENCES

Anicic, D. and Fodor, P. EP-SPARQL: a unified language for event processing and stream reasoning. In *Proceedings of the International World Wide Web Conferences*. Hyderabad, India, pp. 635–644, 2011.

Barbieri, D., Braga, D., and Ceri, S. C-SPARQL: SPARQL for continuous querying. In *Proceedings of the International World Wide Web Conferences*. Madrid, Spain, pp. 1061–1062, 2009.

Barnaghi, P., Presser, M., and Moessner, K. Publishing linked sensor data. In *Proceedings of the International Workshop on Semantic Sensor Networks*. Shanghai, China, pp. 1–16, 2010.

Barrasa, J., Corcho, O., and Gómez-pérez, A. R2O, an Extensible and Semantically based Database-to-Ontology Mapping Language. In *Proceedings of the Workshop on Semantic Web and Databases*. pp. 1069–1070, 2004.

Berners-lee, T. Linked Data - Design Issues. http://www.w3.org/DesignIssues/LinkedData.html., 2006.

Bizer, C. D2R Map: A Database to RDF Mapping Language. In *Proceedings of the International World Wide Web Conferences*. Budapest, Hungary, pp. 2–3, 2003.

Brenninkmeijer, C. and Galpin, I. A semantics for a query language over sensors, streams and relations. In *Proceedings of the British National Conference on Databases: Sharing Data, Information and Knowledge*. Cardiff, Wales, UK, pp. 87–99, 2008.

Calbimonte, J., Jeung, H., Corcho, O., and Aberer, K. Semantic sensor data search in a large-scale federated sensor network. In *Proceedings of the International Workshop on Semantic Sensor Networks*. Bonn, Germany., pp. 23–38, 2011.

Calbimonte, J.-P., Corcho, O., and Gray, A. J. G. Enabling ontology-based access to streaming data sources. In *Proceedings of the International Semantic Web Conference on The Semantic Web*. Shanghai, China, pp. 1–16, 2010.

Caraballo, A., Júnior, N., and Nunes, B. TRTML-A Tripleset Recommendation Tool based on Supervised Learning Algorithms. In *Proceedings of the Extended Semantic Web Conference*. Anissaras, Crete, Greece, pp. 413–417, 2014.

Cheng, B., Longo, S., Cirillo, F., Bauer, M., and Kovacs, E. Building a big data platform for smart cities: Experience and lessons from santander. In *Proceedings of the IEEE International Congress on Big Data*. Washington, DC, USA, pp. 592–599, 2015.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., and Taylor, K. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web* 17 (1): 25–32, 2012.

Compton, M., Henson, C., and Neuhaus, H. A Survey of the Semantic Specification of Sensors. In *Proceedings of the International Conference on Semantic Sensor Networks*. Washington DC, pp. 17–32, 2009.

Consortium, W. W. W. R2RML: RDB to RDF mapping language. http://travesia.mcu.es/portalnb/jspui/handle/10421/7489 http://www.w3.org/TR/r2rml/, 2012.

Cuan, L.-N., Shi, Y.-M., Li, G.-Y., Wu, X.-H., and Liu, N. An algorithm to generate data linkages of linked sensor data. In *Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. pp. 430–433, 2015.

Harth, A., Knoblock, C., and Stadtmüller, S. On-the-fly Integration of Static and Dynamic Linked Data. In *Proceedings of the International Semantic Web Conference*. Sydney, Australia, 2013.

Hasemann, H., Kremer, A., Pagel, M., Group, A., and Braunschweig, T. U. RDF Provisioning for the Internet of Things. In *Proceedings of the International Conference on the Internet of Things*.

Hübner, K., Schünemann, B., and Radusch, I. Sophisticated route calculation approaches for microscopic traffic simulations. In *Proceedings of the International Conference on Simulation Tools and Techniques*. pp. 147–154, 2015.

Le-Phuoc, D., Quoc, H. N. M., Parreira, J. X., and Hauswirth, M. The linked sensor middleware–connecting the real world and the semantic web. Tech. rep., 2011.

Lopes, G., Leme, L., Nunes, B., and Casanova, M. RecLAK: Analysis and Recommendation of Interlinking Datasets. In *Proceedings of the Workshops at the International Conference on Learning Analytics and Knowledge*. Indianapolis, USA, pp. 1–6, 2014.

Lopes, G. R., Andr, L., Leme, P. P., Nunes, B. P., Casanova, M. A., and Dietze, S. Recommending Tripleset Interlinking. In *Proceedings of the International Conference on Web Information System Engineering*. Nanjing, China, pp. 149–161, 2013.

Moraru, A., Fortuna, C., and Mladenic, D. A System for Publishing Sensor Data on the Semantic Web. *Journal of Computing and Information Technology* 19 (4): 239–245, 2011.

Mukhopadhyay, S. C. Wearable sensors for human activity monitoring: A review. *IEEE Sensors Journal* 15 (3): 1321–1330, 2015.

Patni, H., Henson, C., and Sheth, A. Linked sensor data. In *International Symposium on Collaborative Technologies and Systems*. pp. 362–370, 2010.

Phuoc, D. and Hauswirth, M. Linked open data in sensor data mashups. *In proceedings of the International Workshop on Semantic Sensor Networks*, 2009.

Phuoc, D. L. *A Native and Adaptive Approach for Linked Stream Data Processing*. Ph.D. thesis, National University of Ireland, Ireland, 2013.

Rezvan, M., Barekatain, M., Zaeri, A., and Taghandiki, K. Applying an innovative semantic sensor network model in internet of things. In *Proceedings of the International Conference on Information and Communication Technology Convergence*. pp. 324–328, 2015.

Sheth, A., Henson, C., and Sahoo, S. Semantic sensor web. *Internet Computing* 12 (4): 78–83, 2008.

Soldatos, J., Kefalakis, N., Serrano, M., and Hauswirth, M. Design principles for utility-driven services and cloud-based computing modelling for the internet of things. *International Journal of Web Grid Service* 10 (2/3): 139–167, 2014.

Yu, L. and Liu, Y. Using linked data in a heterogeneous sensor web: challenges, experiments and lessons learned. *International Journal of Digital Earth* 8 (1): 17–37, 2015.

Zhang, Q., Zhang, Y., and Li, J. Easycomeeasygo: Predicting bus arrival time with smart phone. In *Proceedings of the International Conference on Frontier of Computer Science and Technology*. pp. 268–273, 2015.

Zhang, X., Zhao, Y., and Liu, W. A method for mapping sensor data to ssn ontology. *International Journal of u-and e-Service, Science and Technology* 8 (9): 303–316, 2015.