6

A Review of Tools for IoT Semantics and Data Streaming Analytics

Martin Serrano and Amelie Gyrard

Insight Center for Data Analytics, National University of Ireland, Galway, Ireland

6.1 Introduction

The Internet of Things is having a fast adoption in the industry and with no doubt, in today's increasingly competitive innovation-driven marketplace, individuals need skills to transform the growing amount of industry, product, and customer (behavior) data into actionable information to support strategic and tactical decisions at the organisations. The Internet of Things (Devices) and the web are every day more and more crucial elements in this data generation, consolidating the need for interpreting data produced by those devices or "things", physical or virtual, connected to the Internet. In the other hand analytics is a time-consuming process and constantly redesigned in every application domain. The identification of methods and tools to avoid that data analytics over IoT Data is every time re-designed is a constant need. In this chapter, a review of tools for IoT data analytics is reviewed. We provide an overall vision on best practices and trends to perform analytics, we address innovative approaches using semantics to facilitate the integration of different sources of information [1].

The Semantic Web community plays a relevant role when interoperability of data and integration of results are required. The semantic analytics is an emerging initiative talking about Linked Open Data (LOD) reasoning, provides a vision on how to deduce meaningful information from IoT data, aiming to share the way to interpret data in a interoperable way to produce new knowledge [2]. Semantic analytics unifies different semantics technologies and analytic tools such as logic-based reasoning, machine learning, Linked Open Data (LOD), the main objective is to convert data into actionable knowledge, reasoning combines both semantic web technologies and reasoning approaches.

Collecting, transforming, interpreting data produced by devices "things" connected to the Internet/Web is a time-consuming process and constantly requires a redesign process for all applications that uses different data sources [3].

In this chapter, we analyse complementary research fields covering reasoning and semantic web approaches towards the common goal of enriching data within IoT. The involvement of semantics offers new opportunities and methods for production and discovery of information and also transforms information into actionable knowledge. The most common used methods are: 1) Linking Data, 2) Real-time and Linked Stream Processing, 3) Logic-based approaches, 4) Machine Learning based approaches, 5) Semantics-based distributed reasoning, and 6) Cross-domain recommender systems.

Ozpinar in 2014 [4] explained that resolving the meaning of data is a challenging problem and without processing it the data is invaluable. Pereira in 2014 [5] highlighted the necessity to interpret, analyse and understand sensor data to perform machine-to-machine communications. They classify six techniques such as supervised learning, unsupervised learning, rules, fuzzy logic, ontological reasoning and probabilistic reasoning in their survey dedicated to context-awareness for IoT. Further, they clearly explain pros and cons and sum up them in a table. According to their table, rule and ontology-based techniques contain few cons. Their shortcomings are to define manually rules which can be error prone and that there is no validation or quality checking. With such approaches, rules are only defined once in an interoperable manner. Pros concerning rule-based system are that rules are simple to define, easy to extend and require less computational resources. In semantic analytics and particularly 'Sensor-based Linked Open Rules' [6] will overcome these limitations, rules can be shared and reused and validated by domain experts. To deduce meaningful information from sensor data, the following main challenges to address are analysed: a) Real-time data, b) Scalability, c) Which machine learning algorithm should be apply for specific sensor datasets because there is a need to assist users in choosing the algorithm fitting their need, e) How to unify exiting systems and tools (e.g, S-LOR, LD4Sensors, KAT and LSM) since they are providing complementing approaches towards the same goal of enriching data, and f) How to extend KAT to assist experimenters to deal with machine learning and with real-time and to be compatible with the Stream Annotation Ontology (SAO).

Currently what is missing, are the methods to design innovative approaches for linked open data analytics. Recent approaches like Linked Open Reasoning (LOR), introduced in [2], inspired from the recent work in the context of European research projects. Linked Open Reasoning provides a solution to deduce meaningful information from IoT data and aims to share the way to interpret data in an interoperable way. This approach unifies different reasoning approaches such as logic-based reasoning and machine learning. Linked Open Reasoning combines both semantic web technologies and reasoning approaches. There is a vision that machine-learning approaches might not be necessary to interpret data produced by simple sensors. It will avoid the learning curve to deal with machine learning algorithms. The idea of "Linked Open Reasoning" (LOR) is an extension of our preliminary idea, Sensor-based Linked Open Rules (S-LOR) [6]. S-LOR is a dataset of interoperable IF THEN ELSE rules to deduce meaningful information from simple sensors such as thermometer.

6.2 Related Work

In this article, a complementary research covering reasoning and semantic web approaches towards the common goal of enriching data within IoT is presented and studied following the different approaches: 1) Linking Data, 2) Real-time and Linked Stream Processing, 3) Logic-based approaches, 4) Machine Learning based approaches, 5) Semantics-based distributed reasoning, and 6) Cross-domain recommender systems. We conclude by comparing different approaches and tools and highlighting the main limitations.

6.2.1 Linking Data

Karma is a data integration tool dealing with heterogeneous data such as XML, CSV, JSON, Web APIs, etc. based on ontologies and eases the publication of data semantically annotated with RDF [7]. This tool has been used to aggregate smart city data and semantically annotate data according to the KM4City ontology [8].

The LD4Sensors/inContext-Sensing is a tool that has been designed within the SPITFIRE EU project. This tool enriches sensor data with the Linked Data by using the Pachube API, the SPITFIRE ontology and the Silk tool to align datasets such as DBPedia, WordNet, Musicbrainz, DBLP, flickr wrappr and Geonames [9]. LD4Sensors provides JSON Web services, API and GUI to automate the annotation and linking process of sensor data. The semantic annotation is done with the Jena library and semantic data are stored using Jena TDB. LD4Sensors integrates a SPARQL endpoint to ease access to semantic sensor data. The semantic dataset and SPARQL endpoint are referenced on the dataset catalogue called DataHub. LD4Sensors provide linking but do not deal with real-time aspect nor interpret sensor data produced by devices by reusing domain-specific knowledge expertise.

6.2.2 Real-time & Linked Stream Processing

In recent years, a significant number of technologies that facilitate real-time and linked stream processing have also emerged. **Linked Stream Data** is an extension of the SPARQL query language and engine to deal with stream sensor data and enrich them with the Linked Open Data cloud [10]. **SPARQL is an RDF query language** and protocol produced by the W3C RDF Data Access Working Group (DAWG). SPARQL is extensively used in Semantic communities and was released as a W3C Recommendation in 2008. **C-SPARQL** was an earlier proposal of streaming SPARQL system [11]. Furthermore, the Continuous Query Evaluation over Linked Streams (CQELS) combines streaming capabilities and Linked Data [12, 13]. On top of this Le-Phuoc et al. [14] developed the SensorMasher and Linked Sensor Middleware (LSM) platforms in order to facilitate publishing of 'Linked Stream Data' and their use within other applications. In particular, they developed a user friendly interface to manage environmental semantic sensor networks. SPARQLStream is another novel approach for accessing and querying existing streaming data sources [15]. Specifically, SPARQLstream has been designed as an extension to the SPARQL 1.1 query language to deal with real-time sensor data [16].

6.2.3 Logic

Several mechanism and tools have also been developed in order to apply processing logic over streams. For example, Sensor-based Linked Open Rules (S-LOR) is an approach to share and reuse the rules to interpret IoT data, as explained in Section 6.3.3. It provides interoperable datasets of rules compliant with the Jena framework and inference engine. The rules have been written manually but are extracted from the Linked Open Vocabularies for Internet of Things (LOV4IoT) dataset [62, 63], an ontology/dataset/rule catalogue designed by domain experts in various applicative domains relevant for IoT such as healthcare, agriculture, smart home, smart city, etc.

Linked Edit Rules (LER) [17] is another recent approach similar to the Sensor-based Linked Open Rules (S-LOR) to share and reuse the rules associated to the data. This work has been not applied to the context of IoT. LER is more focused on checking consistency of data (e.g., a person's age cannot be negative, a man cannot be pregnant and an underage person cannot process a driving license). LER extends the RDF Data Cube data model by introducing the concept of EditRule. The implementation of LER is based on Stardogs rule reasoning to check obvious consistency.

Another relevant approach is provided by the **BASIL framework (Building APIs SImpLy)**, which combines REST principles and SPARQL endpoints in order to benefit from Web APIs and Linked Data [18]. BASIL reduces the learning curve of data consumers since they query web services exploiting SPARQL endpoints. The main benefit is that data consumers do not need to learn the SPARQL language and semantic web technologies.

6.2.4 Machine Learning

Machine learning is one of the most extended techniques in information systems, [19] and [20] are the earlier work to propose the idea to reason on semantic sensor data (e.g., to deduce potentially icy, blizzard, freezing concepts). The work described in [21] explains the idea of 'semantic perception' [22, 23] to interpret and reason on sensor data. This work developed an ontology of perception called IntellegO. A semantic-based approach to integrate abductive logic framework and Parsimonious Covering Theory (PCT) to integrate semantics in resource-constrained devices was also proposed. It explains that the development of background knowledge is a difficult task and out of the scope of this work. For this reason, recently, the LOV4IoT dataset has been designed to encourage the reuse of the domain knowledge expertise relevant for IoT. LOVIoT shows numerous challenges to automatically combine the background knowledge. IntellegO also illustrates that perception does not enable a straightforward formalization using logic-based reasoning. e.g., for simple sensors such as temperature or precipitation, logic-based reasoning is faster, flexible and easier for sharing. For more complex sensors such as accelerometers or ECG, logic-based reasoning is insufficient, and the uses of data mining approaches are unavoidable.

Beyond learning about the data, there are work (i.e. [24, 25]) introducing a Knowledge Acquisition Toolkit (KAT) to infer high-level abstractions from sensor data provided by gateways in order to reduce the traffic in network communications. KAT comprises three components: 1) An extension of Symbolic Aggregate Approximation (SAX) algorithm, called SensorSAX, 2) Abductive reasoning based on the Parsimonious Covering Theory (PCT), and 3) Temporal and spatial reasoning. It uses machine learning techniques (i.e. k-means clustering and Markov model methods) and Semantic Web Rule Language (SWRL) rule-based systems to add labels to the abstractions. KAT proposes the use of domain-specific background knowledge, which is not sufficient for Internet of Things, unless some another approach (e.g., the LOV4IoT dataset) is also integrated.

The work described in [26] employs the abductive model rather than inductive or deductive approaches to solve the incompleteness limitation due to missing observation information. The work is tested on real sensor data (i.e. temperature, light, sound, presence and power consumption). Their gateways support TinyOS, Contiki enabled devices and Oracle SunSpot nodes.

There are also approaches that emphasize the use of machine learning on sensor data [27]. This includes for example the use of decision trees and Bayesian network to analyze datasets comprising 16,578 measurements. The focus of the approach is on four kinds of sensor measurements: temperature, humidity, light and pressure. Furthermore, the dataset used has additional information such as weekday, hour interval, position of the window, number of computers working and number of people in the lab. An enrichment of sensor data with semantics has been also taken place [28, 29]. This enrichment provides context for sensor measurements, based on well-known ontologies such as Geonames for location, Geo WGS84 for coordinates, the W3C SSN ontology to describe sensors, the SWEET ontologies, as well as the W3C Time ontology. However, no need for IoT-related domain ontologies is expressed, while the need for semantic reasoners as a mean to infer new knowledge from sensor data is outlined [28].

The SemSense architecture [30] is one more approach to collect and then publish sensor data as Linked Data. Also, in [31], the authors collect data on the fly and then validate and link them with Linked Open Data (LOD) datasets. Devaraju et al. have also designed an ontology for weather events observed by sensors such as wind speed and visibility [32]. They are focused on blizzard related phenomena. They deduce high-level abstractions such as the types of snow (e.g., soft hail, snow, snow pellet, blizzard, winter storm, avalanche, flood, drought, and tornado). Such abstractions are deduced with rule-based reasoning, the implementation is based on Semantic Web Rule Language (SWRL)[1] and the Jess reasoning engine. The DUL ontology and the W3C SSN ontology are used. The approach is evaluated based on the Canadian Climate Archives database. In another work, Wang et al. explain that the SSN ontology "does not include modeling aspects for features of interest, units of measurement and domain knowledge that need to be associated with the

[1]https://www.w3.org/Submission/SWRL/

sensor data to support autonomous data communication, efficient reasoning and decision making" [33].

In recent years, there have also been efforts to interpret data produced by accelerometer, gyroscope, microphone, temperature and light sensors embedded in mobile phones [34]. These efforts use Hidden Markov Models (HMMs) and semantic web technologies to deduce activities. The rules are implemented as SPARQL queries. Moreover, Ramparany et al. introduced the need of a domain-specific automated reasoning system [35]. This work envisages that such a system could be based on Description Logic or Complex Event Processing (CEP) for interpreting IoT data. However, it does not propose a dataset with predefined rules that could be easily shared and reused by developers.

6.2.5 Semantic-based Distributed Reasoning

One of the early work on semantic-based distributed reasoning has been DRAGO, the Distributed Reasoning Architecture for a Galaxy of Ontologies, implemented as a peer-to-peer architecture [36]. The goal of DRAGO was to reason on distributed ontologies. Likewise, Kaonp2p has been designed to query over distributed ontologies [35]. Moreover, LarKC (Large Knowledge Collider) is another scalable platform for distributed reasoning [37]. Similarly, the Marvin framework is a scalable platform for parallel and distributing reasoning on RDF data [38]. Also, Schlicht et al. propose a peer-to-peer reasoning for interlinking ontologies [36]. These works outline the need to provide interoperable heterogeneous sensor-based rules and combine cross-domain ontologies and datasets in the context of IoT applications.

Abiteboul et al. have also approached the Web as a distributed knowledge base and proposed an automated reasoning over it [40]. This work demonstrated the importance of reusing sensor-based domain ontologies and rules. Also, WebPIE (Web-scale Parallel Inference Engine) is an inference engine for semantic web reasoning (OWL and RDFS) based on the Hadoop BigData platform [41]. WebPIE is scalable over 100 billion triples [42]. Another scalable system has been introduced by Coppens et al. [43] as an extension to the SPARQL query language to support distributed and remote reasoning. The relevant implementation of the system has been based on the Jena ARQ query engine. One more semantic reasoning framework for BigData has been introduced by Park et al. based on XOntology and SPARQL [44]. It uses the Hadoop platform, HDFS and MapReduce to deal with thousands of sensor data nodes.

Overall, it is noteworthy that none of the discussed distributed reasoning frameworks proposes and implements interoperable rules as a means of interpreting sensor data.

6.2.6 Cross-Domain Recommender Systems

Recently, cross-domain semantic and rule-based recommender systems have also been designed [45] or [46]. Such systems underline the importance of providing interoperable reasoning. Some works (e.g., [47, 48]) propose a domain-independent recommendation system to provide personalization services of different domains (tourism, movies, books). They incorporate semantics into a content-based system to improve the flexibility and the quality, a domain-based inference (side-ward propagation, upward propagation) for user's interests and a semantic similarity method is used to refine item-user matching algorithm. Such cross-domain recommender systems highlight the importance to provide a domain-independent reasoning.

At the same time, Hoxha et al. provide a cross-domain recommender system based on semantics and machine learning techniques (Markov logic) [45], while Tobias et al. provide a context-aware cross-domain recommender system. They exploit semantic web technologies and related tools such as DBpedia and the spreading activation algorithm [46]. These works underline the importance of a cross-domain reasoning that could also applied to sensor data.

6.2.7 Limitations of Existing Work

Most of the presented works have limitations when it comes to adding semantic capabilities for analytics in an IoT context. For example:

- LD4Sensors does not deal with real-time data and does not provide inference reasoning to deduce new information. However, datasets have been linked to get additional information.
- LSM does not integrate inference-reasoning engine to deduce new information.
- KAT has some usability limitations. Non-experts in machine learning have some difficulties to use this tool since they have to choose the algorithm without any assistance.
- S-LOR provides interoperable Jena rules. However, the same rules could be designed with SPARQL CONSTRUCT. Since SPARQL is a recommendation it would be better to share and reusing the rules according to SPARQL CONSTRUCT.

In Figure 6.1, we indicate the pros and cons of different approaches to enrich IoT data on the basis of: (A) Logic or rule-based reasoning, (B) Machine learning, (C) Linked Stream processing, (D) Reusing domain knowledge with

Methods	Pros	Cons
Logic/rule-based reasoning	- Simple rules - Adapted to simple sensors - Easy for beginners (learning & implementation) - Easier to combine rules	- Not adapted to complicated sensors - Heterogeneous rule languages and editors
Machine learning	- More elaborate results - Adapted to complicated sensors	- Need real datasets - Complicated for non-experts - Complicated for a 'sharing and reusing' approach
Linked Stream processing	- Real-time data - Scalability - Linked Data	- No real reasoning
Re-use domain knowledge (LOD, LOV, LOR)	- 'Sharing and reusing' approach	- Be familiar with semantics - Be familiar with ontology/instance matching tools - Not adapted to complicated sensors
Distributed Reasoning	- Scalability - Interoperability between systems	- Complicated for implementation
Recommendation systems	- Adapted to the user profile	- Complicated for non-experts - Need real datasets - Need user profile

Figure 6.1 Summary of existing approaches for IoT data enrichment.

Linked Open Data (LOD), Linked Open Vocabularies (LOV) and Linked Open Rules (LOR); (E) Distributed reasoning, and (F) Recommender systems.

Figure 6.2 depicts a classification of different tools according to the different approaches. Some of the existing works are based on machine learning algorithms. Usually, machine learning is employed when rule-based algorithms are infeasible. None of the existing works deals with the extraction, reuse and linking of rules already implemented in domain-specific projects. To deal with such limitations, there is a need to build a dataset of interoperable rules to reason on sensor data. To achieve this task, sensor data should be interoperable. This approach should be easy to be shared and reused by other projects. Since, SWRL rules are increasingly popular the approach will be based on this language. Further, sharing, reusing and combining SWRL rules will be typically easier than data mining algorithms.

6.3 Semantic Analytics

The semantic web community has designed open approaches for sharing and reusing open data by means of using Linked Data, Linked Vocabularies, and Linked Services as a first approach for enabling analytics. Inspired from

Methods / Tools	Logic/rule-based reasoning	Machine learning	Linked Stream processing	Re-use domain knowledge (LOD, LOV, LOR)	Distributed Reasoning	Recommendation systems
'Semantic Sensor Web'	Yes	No	No	Yes	No	No
'Semantic Perception'	No	Yes	No	No	No	No
KAT	No	Yes	No	No	No	No
inContext-sensing	No	No	No	Yes	No	No
S-LOR	Yes	No	No	Yes	No	No
CQELS	No	No	Yes	Yes	No	No
SPARQLStream	No	No	Yes	Yes	No	No
WebPIE	No	No	No	No	Yes	No
DRAGO	No	No	No	No	Yes	No
Marvin	No	No	No	No	Yes	No
KaonP2P	No	No	No	No	Yes	No
LarKC	No	No	No	No	Yes	No

Figure 6.2 Classification of tools according to reasoning approaches.

the Semantic Web community semantic analytics plays a relevant role when interoperability of data and integration of results are required.

6.3.1 Architecture towards the Linked Open Reasoning

In this section, we describe and we attempt to unify reasoning approaches. Figure 6.3 presents a summary of the studied reasoning tools divided, for an easier understanding, in 3 layers and summarised as follow:

- The first layer (top) shown at the bottom provides API and web services to access to reasoning approaches. This layer provides access to simple reasoning services or complex services which are a composition of existing services.
- The second layer (middle) shows the generic reasoning approaches. We referenced, classified and analyzed the following different reasoning methods, including: (A) Machine Learning, which is a quite popular approach, yet it needs data to be integrated, which are not always available; (B) Real-time techniques, which are important when dealing with real-time data; (C) Linking techniques, which enable the enrichment of IoT data with background knowledge; (D) Complex Event Processing (CEP) approaches, which apply inference based reasoning in order to extract or deduce new information from IoT data; (E) Sharing and

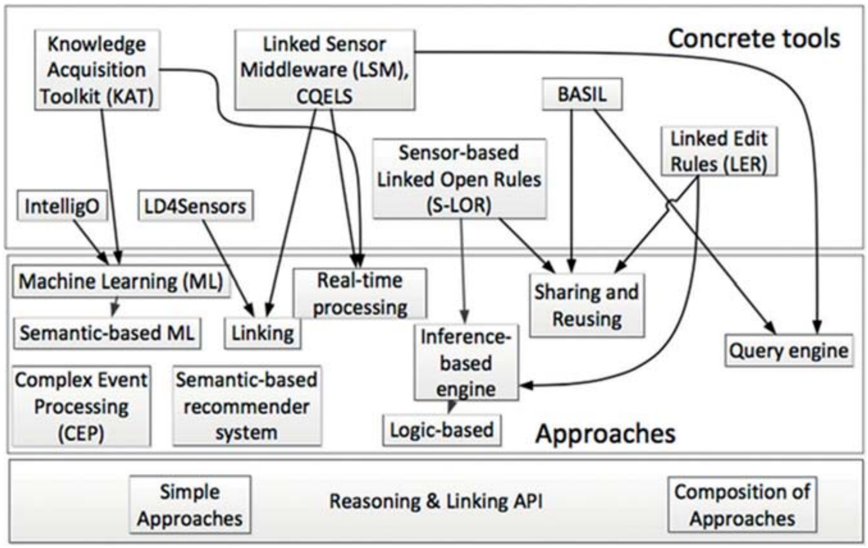

Figure 6.3 Reasoning main operations.

reusing approaches, which enable deduction of new information from data produced by Internet Connected Objects (ICOs) in a way similar to Linked Open Data.

- The third layer (bottom) indicates the concrete tools that we could reuse and unify to interpret data. Such tools are: (A) the Knowledge Acquisition Toolkit (KAT) which is a machine-learning approach dealing with real-time data; (B) The Linked Sensor Middleware (LSM), which deals with real-time data and enables linking between heterogeneous datasets; (C) IntelligO, which is a machine-learning approach using the Parsimonious Covering Theory (PCT); (D) Linked Data for Sensors (LD4Sensors), which enables linking of sensor datasets and (E) Sensor-based Linked Open Rules (S-LOR) which is a rule-based reasoning engine and innovative approach to share and reuse interoperable deductive rules in order to infer new knowledge from IoT data.

6.3.2 The Workflow to Process IoT Data

Figure 6.4 illustrates different processes and steps required to combine data from heterogeneous sources and to build innovative and interoperable applications. The figure illustrates the SEG 3.0 methodology [2] that is extended and used for interoperability but also for semantic analytics. In this book chapter, we are mainly focused on the reasoning layer. It comprises the following steps:

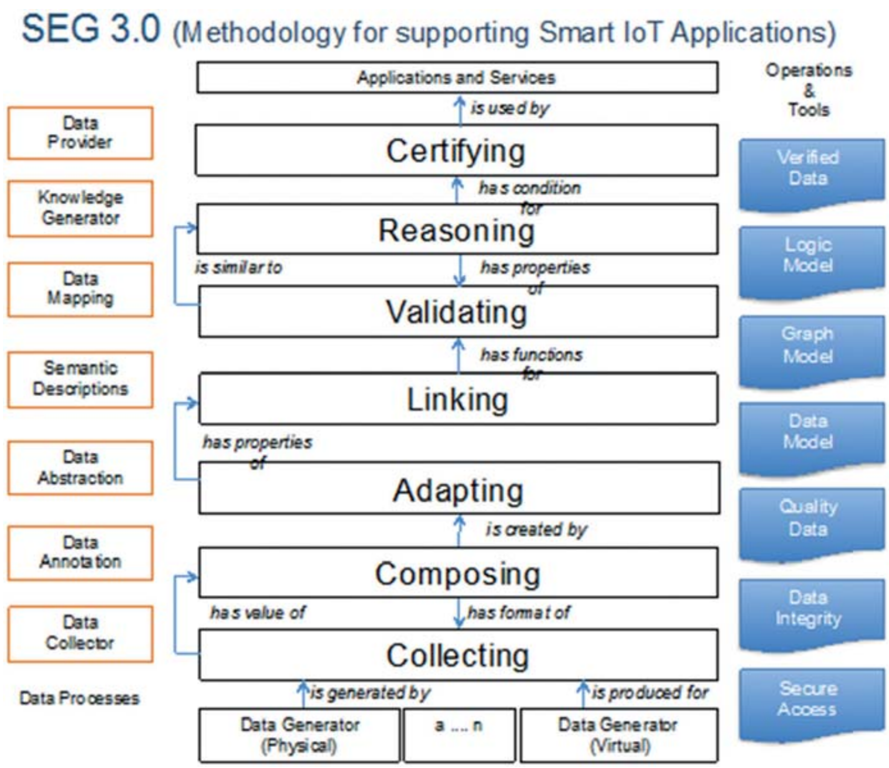

Figure 6.4 IoT process defined by SEG 3.0 methodology.

- **Composition of IoT data**, enabling the unification of heterogeneous data coming from different IoT sources and particularly re-using different data formats (e.g., CSV, Excel) or different terms (e.g., temp or temperature). This activity requires a common dictionary to unify terms employed to describe data. For example, the composing layer could return the SenML format to describe sensor data [49].
- **IoT Data Modeling**, which enables the annotation of data with semantic web technologies (e.g., RDF, RDFS and OWL). This step employs models, vocabularies and ontologies to unify data, which is prerequisite for the next steps. The M3 ontology is used to unify semantic sensor data [50].
- **Linking IoT data domains**, enabling the enrichment of data with metadata from other RDF datasets to get additional information. It exploits the idea of Linked Data and Linked Vocabularies for IoT applications.

- **Reasoning over IoT Streaming Data**, which enables the updating of the database/triple store with additional triples for instance by using a reasoning engine (e.g., Jena rule-based inference engine) to infer high level abstraction from data. It exploits the idea of Linked Rules.
- **Querying IoT Data**, which enables the querying of RDF datasets through the SPARQL language based on ontologies used in the previous steps. It is an essential step to get data and build end-users services/applications.
- **IoT Services activation and control**, which enables end-users to access smarter data. The data is available through interoperable APIs or web services (e.g., RESTful web services). Such web services returns the result provided by the SPARQL query engine.
- **Composition of IoT services**, which enables the development of complex applications based on the composition of several services. It can be achieved through the use of web services or semantic web services.

The SEG 3.0 methodology supports the vision of semantic interoperability from data to end-users applications, which is inspired from the 'sharing and reusing' based approach as depicted in Figure 6.4. The realization of the vision is based on a combination of several of the concepts that have been already presented including: 1) Linked Open Data (LOD); 2) Linked Open Vocabularies (LOV); 3) Linked Open Rules/Reasoning (LOR); and 4) Linked Open Services (LOS).

In the following paragraphs, we extend and apply these approaches to IoT and smart cities. Note that Linked Open Data (LOD) is an approach to share and reuse the data [51, 52]. However, previous works on 'Linked Sensor Data' [53, 54] do not provide any tools for visualizing or navigating through IoT datasets. For this reason, a Linked Open Data Cloud for Internet of Things (CLOuDIoT) infrastructure to share and reuse data produced by sensors is being implemented.

Linked Open Vocabularies (LOV) is an approach to share and reuse the models/vocabularies/ontologies [55]. LOV did not reference any IoT ontologies. For this reason, we have also designed the Linked Open Vocabularies for Internet of Things (LOV4IoT) [62, 63], a dataset of almost 300 ontology-based IoT projects referencing and classifying: 1) IoT applicative domains, 2) Sensors used, 3) Ontology status (e.g., shared online, best practices followed), 4) Reasoning used to infer high level abstraction, and 5) Research articles related to the project. This dataset contains the background knowledge required to add value to the data produced by Internet Connected Objects (ICOs).

6.3.3 Sensor-based Linked Open Rules (S-LOR)

Linked Open Reasoning (LOR) is an approach for sharing and reusing the ways to interpret the data and to deduce new information (e.g., machine learning algorithm used, reusing rules already designed by domain experts). To this end, LOR can be extended towards using semantics Sensor-based Linked Open Rules (S-LOR), a dataset of interoperable rules (e.g., if-then-else rules) used to interpret data produced by sensors [56]. Such rules are executed with an inference engine (e.g., Jena) that updates the triple store with additional triples. For example, the rule can be if the body temperature is greater than 38 degree Celsius than fever. In this example, the triple store will be updated with this high level abstraction 'fever'. The approach is inspired from the idea of 'Linked Rules' [57] that provides a language to interchange semantic rules but not the idea of reusing existing rules.

6.4 Tools & Platforms

Data analytics is a complex activity that requires examining raw data with the purpose of drawing conclusions. In IoT, the combination of different data types, in nature and in format, makes this practice more complex. Data analytics is extensively used in science to verify or eliminate existing models or theories, analytics is also used in many domains to allow companies and organization to make better business decisions. Data analytics focuses on inference, the process of generate conclusion(s) based solely on what is already known by the researcher.

On the other hand, semantic analytics is an advanced technique that uses the normalisation of data to a one particular format with the advantage of data alignment and interoperability. This allows the generation of more information. Necessary steps for semantic analytics, along with related tools are presented in the following paragraphs.

6.4.1 Semantic Modeling and Validation Tools

A variety of semantic modeling tools have recently emerged and are already used in the scope of IoT applications. For example:

- **HyperThing**[2] is a semantic web URI validator, which determines whether a URI identies a Real World Object or a Web document resource. It checks whether the URIs publishing method follows the W3C hash URIs and 303 URI practices. It can also be used to check the validity of the

[2]http://www.hyperthing.org

chains of the redirection between the Real World Object URIs and Document URIs, in order to prevent the data publisher mistakenly redirect.

- **NeON**[3] is a methodology for Ontology Engineering in a networked world.
- **OWL Validator** is another semantic validation project that accepts ontologies written in RDF/XML, OWL/XML, OWL Functional Syntax, Manchester OWL Syntax, OBO Syntax, and KRSS Syntax.
- **OQuaRE**[4] is a square-based approach for evaluating the quality of ontologies. OQuaRE covers two main processes: software quality requirement specifications and software quality evaluation.
- **OntoClean**[5] provides a definition of metaproperties that help with the construction of ontology language descriptions of problem domains.
- **OnToology**[6] is a system to automate part of the collaborative ontology development process. OnToology works surveying a repository with an OWL file, produce diagrams, a complete documentation and do validation based on common pitfalls.
- **Oops**[7] helps ontology designers detect some of the most common pitfalls appearing within ontology developments in particular when: (a) the domain or range of a relationship is defined as the intersection of two or more classes; (b) no naming convention is used within the identifiers of the ontology elements; and (c) a cycle between two classes in the hierarchy is included in the ontology.
- **Ontocheck**[8] is a tool for verifying ontology naming conventions and metadata completeness following cardinality checks on mandatory and obligatory annotation properties and reviewing naming conventions via lexical analysis and labeling enforcement.
- **OntoAPI**[9] is a consumer API that can process the response of an ontology evaluation web service provider.
- **OntoMetric**[10] is a method to choose the appropriate ontology.
- **Prefix**[11] simplifies the tedious task of any RDF developer, by remembering and looking up URI prefixes.

[3] http://neon-toolkit.org/wiki/Main Page
[4] http://miuras.inf.um.es/oquarewiki/index.php5/MainPage
[5] http://www.ontoclean.org/
[6] http://ontoology.linkeddata.es
[7] http://oops.linkeddata.es
[8] http://www2.imbi.uni-freiburg.de/ontology/OntoCheck/
[9] https://sourceforge.net/projects/drontoapi/
[10] http://oa.upm.es/6467/
[11] http://prefix.cc/

- **Vapour**[12] is a linked Data Validator in the form of a scripting approach to debug content. Vapour facilitates the task of testing the results of content negotiation on a vocabulary.
- **Vocab**[13] is an open source project that allows RDF developers to look up and search for Linked Data vocabularies. Developers can search URIs with arbitrary queries or look up specific URIs.
- **The W3C RDF Validator**[14] is an online service for checking and visualizing your RDF documents, W3C RDF validator is based on Another RDF Parser.

6.4.2 Data Reasoning

There are also numerous data reasoners that enable knowledge generation and activation. However not all of them are in a mature stage nor serve the same purpose. By means of its level of complexity to configure actionable data in the IoT, reasoners can be catalogued not only by their linkage and discovery mechanisms, but also on the basis of their usability in the area. The following selection is a comprehensive list of reasoners that can be used in the scope of IoT streaming and IoT analytics applications:

- **CEL DL (Description Logic)**[15] is a reasoner which implements a polynomial-time algorithm. The supported description logic (EL+) offers a selected set of expressive means that are tailored towards the formulation of domain-specific ontologies. CEL's main reasoning task is the computation of the subsumption hierarchy induced by EL+ ontologies.
- **Euler**[16] is an inference engine supporting logic-based proofs. It is a backward-chaining reasoner enhanced with Euler path detection. It has implementations in Java, C#, Python, JavaScript and Prolog. In conjunction with N3 it is interoperable with W3C Cwm.
- **FaCT++**[17] is the new generation of the well-known FaCT OWL-DL reasoner. FaCT++ uses the established FaCT algorithms, but with a different internal architecture.
- **HermiT**[18] is a highly efficient OWL reasoner. HermiT is a reasoner for ontologies written using the Web Ontology Language (OWL). HermiT

[12] http://linkeddata.uriburner.com:8000/vapour
[13] http://vocab.cc
[14] http://www.w3.org/RDF/Validator/
[15] https://lat.inf.tu-dresden.de/systems/cel/
[16] https://www.w3.org/2001/sw/wiki/Euler
[17] http://owl.man.ac.uk/factplusplus/
[18] http://www.hermit-reasoner.com

is based on a novel "hypertableau" calculus which provides much more efficient reasoning than any previously-known algorithm.

- **JESS (Java Expert System Shell)**[19] is in the form of a Jena inference implementation with rule engine and scripting environment written entirely in Java^{TM}.
- **Jena Eyeball**[20] is a command-line semantics validator for checking RDF/OWL common problems. Eyeball is a Jena-based tool for checking RDF models (including OWL) for common problems. It is user-extensible using plugins.
- **Kaon2**[21] is an OntoBroker designed for managing ontologies. KAON2 is a successor to the KAON project often referred to as KAON1. The main difference to KAON1 is the supported ontology language: KAON1 used a proprietary extension of RDFS, whereas KAON2 is based on OWL-DL and F-Logic.
- **Nools**[22] is a RETE based rule engine written entirely in javascript. When using nools tool, a flow which acts as a container for rules that can later be used to get a session.
- **OWLlink API**[23] is designed to access remote reasoners. OWLlink API has a Java interface for the OWLlink protocol on top of the Java-based OWL API. The OWLlink API enables OWL API-based applications to access remote reasoners (so-called OWLlink servers), and it turns any OWL API aware reasoner into an OWLlink server.
- **Pellet**[24] is an open-source Java based OWL 2 reasoner, It can be used in conjunction with both Jena and OWL API libraries.
- **Racer Pro**[25] is an OWL reasoner tool that can perform reasoning tasks. Racer pro has an inference server for the Semantic Web.
- **RIF4j**[26] is a reasoning engine for RIF-BLD that provides a Java object model for RIF-BLD and supports the parsing and serialization of RIF-BLD formulas. Furthermore, it provides a prototype implementation of a RIF-BLD consumer based on the Datalog engine IRIS.

---

[19] http://www.jessrules.com
[20] http://jena.apache.org/documentation/tools/eyeballgetting-started.html
[21] http://kaon2.semanticweb.org
[22] http://c2fo.io/nools
[23] http://owllink-owlapi.sourceforge.net
[24] https://www.w3.org/2001/sw/wiki/Pellet
[25] https://www.w3.org/2001/sw/wiki/RacerPro
[26] http://rif4j.sourceforge.net

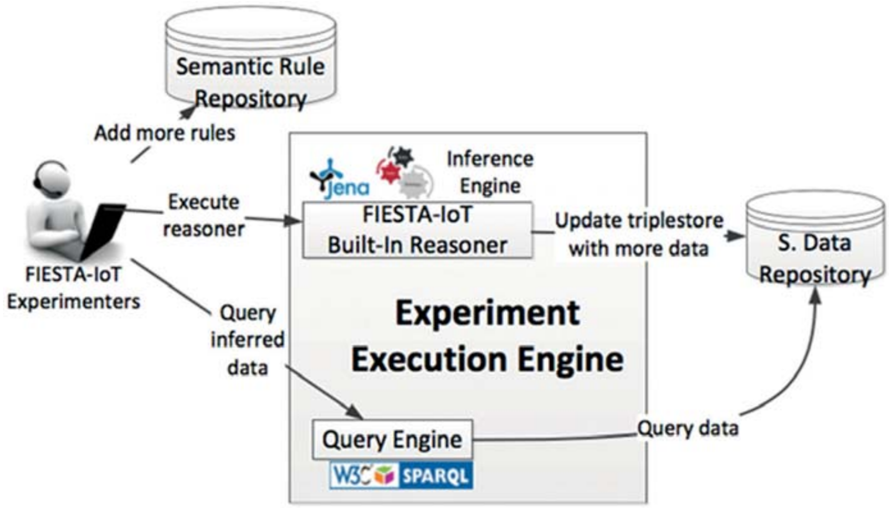

Figure 6.5 IoT reasoning data framework within FIESTA-IoT.

6.5 A Practical Use Case

Federated Interoperable Semantic IoT/cloud Testbeds and Applications (FIESTA-IoT)[27] is an EU project (funded in the context of the H2020 framework), which focuses on integrating IoT platforms, testbeds, data and associated silo applications. FIESTA-IoT opens up new opportunities in the development and deployment of experiments that exploit data and capabilities from multiple geographically and administratively dispersed IT testbeds. The project employs semantic (ontology) modeling as a mechanism to associate different domains and beyond that discover relationships amongst the information.

Figure 6.5 shows the designed FIESTA-IoT reasoning engine approach that by design will be used by experimenters/users of the platform. Based on our analysis of the literature, a logic-based/rule-based reasoning is used. Experimenters can interact with the reasoning engine as follows:

- Increasing the actionable knowledge by contributing to the Semantic Rule Repository, a dataset of interoperable rules. Such rules are IF THEN ELSE rules.

[27]http://fiesta-iot.eu/

- Executing the reasoning engine to infer additional information. Once executed, the reasoning engine updates the triple-store with additional triples (e.g., high level information).
- Querying inferred data, by executing a query engine that interacts with the triple store, called Semantic Data Repository.
- Implementing the rules as Jena rules since we used the Jena framework to build semantic web applications. Moreover, Jena provides an inference engine to easily execute the Jena rules and deduce additional information. To enhance interoperability, Jena rules can be designed as SPARQL CONSTUCT rules.

An overview of the FIESTA-IoT system is provided in Figure 6.5.

6.6 Conclusions

In this chapter, a summary of complementary research fields covering reasoning and semantic web approaches towards the common goal of enriching data within the IoT domain has been presented. The different aspects around semantic analytics like Linking Data, Real-time and Linked Stream Processing, Logic-based approaches, Machine Learning based approaches, semantics-based distributed reasoning, and cross-domain recommender systems, have been summarized and discussed.

The presented approaches and tools are able to deduce meaningful information from IoT data, based on the combination and integration of best practices from the literature. This approach is currently applied in the context of the H2020 FIESTA-IoT project. It leverages a combination of concepts and tools associated with Linked Open Data, Linked Open Vocabularies, Linked Open Services and Linked Open Reasoning.

Acknowledgement

This work is partially funded by the "Federated Interoperable Semantic IoT/cloud Testbeds and Applications" (FIESTA-IoT) project, which is funded by the European Commission in the scope of its Horizon 2020 Programme (Grant Agreement No. CNECT-ICT-643943). The literature survey regarding analytics tools is extended from the authors' past work particularly from the PhD thesis [61] under the supervision of Prof. Christian Bonnet and Dr. Karima Boudaoud.

References

[1] M. Serrano, et al. Cousin, "*Semantic Interoperability: Research Challenges, Best Practices, Solutions and Next Steps*", IERC AC4 Manifesto," 2014.

[2] A. Gyrard and M. Serrano, "*Connected Smart Cities: Interoperability with SEG 3.0 for the Internet of Things*" Proc. 30th IEEE Int'l Conf. Advanced Information Networking and Applications Workshops, 2016.

[3] M. Serrano et al., "*Defining the Stack for Service Delivery Models and Interoperability in the Internet of Things: A Practical Case With OpenIoT-VDK*," IEEE J. Selected Areas in Comm., vol. 33, no. 4, 2015, pp. 676–689.

[4] M. Özpinar, "*A flexible semantic service composition framework for pervasive computing environments*" Ph.D. dissertation, Moddle East Technical University, 2014.

[5] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "*Context aware computing for the internet of things: A survey*", Communications Surveys & Tutorials, IEEE, vol. 16, no. 1, pp. 414–454, 2014.

[6] A. Gyrard, C. Bonnet, and K. Boudaoud, "*Helping IoT application developers with sensor-based linked open rules*," in SSN 2014, 7th International Workshop on Semantic Sensor Networks in conjunction with the 13th International Semantic Web Conference (ISWC 2014), 19–23 October 2014, Riva Del Garda, Italy, 10 2014.

[7] C. A. Knoblock, P. Szekely, J. L. Ambite, A. Goel, S. Gupta, K. Lerman, M. Muslea, M. Taheriyan, and P. Mallick, "Semi-automatically mapping structured sources into the semantic web," in The Semantic Web: Research and Applications. Springer, 2012, pp. 375–390.

[8] P. Bellini, M. Benigni, R. Billero, P. Nesi, and N. Rauch, "Km4city ontology building vs data harvesting and cleaning for smart-city services," Journal of Visual Languages & Computing, vol. 25, no. 6, pp. 827–839, 2014.

[9] M. Leggieri, A. Passant, and M. Hauswirth, "incontext-sensing: Lod augmented sensor data?" in Proceedings of the 10th International Semantic Web Conference (ISWC 2011). Citeseer, 2011.

[10] J. F. Sequeda and O. Corcho, "Linked stream data: A position paper," 2009.

[11] D. F. Barbieri, D. Braga, S. Ceri, E. Della Valle, and M. Grossniklaus, "C-sparql: Sparql for continuous querying," in Proceedings

of the 18th international conference on World wide web. ACM, 2009, pp. 1061–1062.

[12] D. Le Phuoc, "A native and adaptive approach for linked stream data processing," Ph.D. dissertation, 2013.

[13] D. Le-Phuoc, M. Dao-Tran, J. X. Parreira, and M. Hauswirth, "A native and adaptive approach for unified processing of linked streams and linked data," in The Semantic Web-ISWC 2011. Springer, 2011, pp. 370–388.

[14] D. Phuoc and M. Hauswirth, "Linked open data in sensor data mashups," 2009.

[15] J.-P. Calbimonte, O. Corcho, and A. J. Gray, "Enabling ontology-based access to streaming data sources," in The Semantic Web-ISWC 2010. Springer, 2010, pp. 96–111.

[16] J.-P. Calbimonte, "Ontology-based access to sensor data streams," Ph.D. dissertation, Informatica, 2013.

[17] A. Meroño-Peñuela, C. Guéret, and S. Schlobach, "Linked edit rules: A web friendly way of checking quality of rdf data cubes."

[18] E. Daga, L. Panziera, and C. Pedrinaci, "A basilar approach for building web apis on top of sparql endpoints," 2015.

[19] A. Sheth, C. Henson, and S. Sahoo, "Semantic sensor web," Internet Computing, IEEE, vol. 12, no. 4, pp. 78–83, 2008.

[20] W. Wei and P. Barnaghi, "Semantic annotation and reasoning for sensor data," Smart Sensing and Context, pp. 66–76, 2009.

[21] C. A. Henson, "A semantics-based approach to machine perception," Ph.D. dissertation, Wright State University, 2013.

[22] C. Henson, A. Sheth, and K. Thirunarayan, "Semantic perception: Converting sensory observations to abstractions," Internet Computing, IEEE, vol. 16, no. 2, pp. 26–34, 2012.

[23] P. Barnaghi, F. Ganz, C. Henson, and A. Sheth, "Computing perception from sensor data," 2012.

[24] F. Ganz, P. Barnaghi, and F. Carrez, "Information abstraction for heterogeneous real world internet data," Sensors Journal, IEEE, vol. 13, no. 10, pp. 3793–3805, Oct 2013.

[25] F. Ganz, P. Barnaghi, and F. Carrez, "Automated semantic knowledge acquisition from sensor data," 2014.

[26] F. Ganz, "Intelligent communication and information processing for cyber-physical data," Ph.D. dissertation, University of Surrey, 04 2014.

[27] A. Moraru, M. Pesko, M. Porcius, C. Fortuna, and D. Mladenic, "Using machine learning on sensor data," CIT. Journal of Computing and Information Technology, vol. 18, no. 4, pp. 341–347, 2010.

[28] A. Moraru, "Enrichment of sensor descriptions and measuremebts using semantic technologies," Master's thesis, 2011.

[29] A. Moraru and D. Mladenić, "A framework for semantic enrichment of sensor data," CIT. Journal of Computing and Information Technology, vol. 20, no. 3, pp. 167–173, 2012.

[30] A. Moraru, D. Mladenic, M. Vucnik, M. Porcius, C. Fortuna, and M. Mohorcic, "Exposing real world information for the web of things," in Proceedings of the 8th International Workshop on Information Integration on the Web: in conjunction with WWW 2011. ACM, 2011, p. 6.

[31] L. Bradeško, A. Moraru, B. Fortuna, C. Fortuna, and D. Mladenić, "A framework for acquiring semantic sensor descriptions (short paper)," Semantic Sensor Networks, p. 97, 2012.

[32] A. Devaraju and T. Kauppinen, "Sensors tell more than they sense: Modeling and reasoning about sensor observations for understanding weather events," International Journal of Sensors Wireless Communications and Control, vol. 2, no. 1, pp. 14–26, 2012.

[33] W. Wang, S. De, G. Cassar, and K. Moessner, "Knowledge representation in the internet of things: semantic modelling and its applications," Automatika-Journal for Control, Measurement, Electronics, Computing and Communications, vol. 54, no. 4, 2013.

[34] B. Boshoven and P. van Bommel, "Personalized life reasoning," Master's thesis, 2014.

[35] F. Ramparany, F. G. Marquez, J. Soriano, and T. Elsaleh, "Handling smart environment devices, data and services at the semantic level with the fi-ware core platform," in BigData (Big Data), 2014 IEEE International Conference on. IEEE, 2014, pp. 14–20.

[36] L. Serafini and A. Tamilin, "Drago: Distributed reasoning architecture for the semantic web," in The Semantic Web: Research and Applications. Springer, 2005, pp. 361–376.

[37] D. Fensel, F. van Harmelen, B. Andersson, P. Brennan, H. Cunningham, E. Della Valle, F. Fischer, Z. Huang, A. Kiryakov, T.-I. Lee et al., "Towards larkc: a platform for web-scale reasoning," in Semantic Computing, 2008 IEEE International Conference on. IEEE, 2008, pp. 524–529.

[38] E. Oren, S. Kotoulas, G. Anadiotis, R. Siebes, A. ten Teije, and F. van Harmelen, "Marvin: Distributed reasoning over large-scale semantic web data," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, no. 4, pp. 305–316, 2009.

[39] A. Schlicht and H. Stuckenschmidt, “Peer-to-peer reasoning for interlinked ontologies,” International Journal of Semantic Computing, vol. 4, no. 01, pp. 27–58, 2010.
[40] S. Abiteboul, E. Antoine, and J. Stoyanovich, “Viewing the web as a distributed knowledge base,” in Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE, 2012, pp. 1–4.
[41] J. Urbani, S. Kotoulas, J. Maassen, F. Van Harmelen, and H. Bal, “Webpie: A web-scale parallel inference engine using mapreduce,” Web Semantics: Science, Services and Agents on the World Wide Web, vol. 10, pp. 59–75, 2012.
[42] J. Urbani, “On web-scale reasoning,” Ph.D. dissertation, 2013.
[43] S. Coppens, M. Vander Sande, R. Verborgh, E. Mannens, and R. Van de Walle, “Reasoning over sparql,” in Proceedings of the 6th Workshop on Linked Data on the Web, 2013.
[44] K. Park, Y. Kim, and J. Chang, “Semantic reasoning with contextual ontologies on sensor cloud environment,” International Journal of Distributed Sensor Networks, vol. 2014, 2014.
[45] J. Hoxha, “Cross-domain recommendations based on semantically-enhanced user web behavior,” Ph.D. dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), Diss., 2014, 2014.
[46] I. F. Tobías, “A semantic-based framework for building cross-domain networks: Application to item recommendation,” Ph.D. dissertation, 2013.
[47] V. Codina and L. Ceccaroni, “Taking advantage of semantics in recommendation systems,” in Proceedings of the 2010 Conference on Artificial Intelligence Research and Development: Proceedings of the 13th International Conference of the Catalan Association for Artificial Intelligence. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2010, pp. 163–172. [Online]. Available: http://dl.acm.org/citation.cfm?id=1893268.1893291
[48] V. Codina and L. Ceccaroni, “A recommendation system for the semantic web,” in Distributed Computing and Artificial Intelligence. Springer, 2010, pp. 45–52.
[49] J. Arkko, “Network working group c. jennings internet-draft cisco intended status: Standards track z. shelby expires: January 18, 2013 sensinode,” 2012, media Type for Sensor Markup Language (SENML), draft-jennings-senml-09 (work in progress).
[50] A. Gyrard and C. Bonnet, “A unified language to describe M2M/IoT data,” M3 2015.

[51] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," International Journal on Semantic Web and Information Systems (IJSWIS), vol. 5, no. 3, pp. 1–22, 2009, http://www.w3.org/DesignIssues/LinkedData.html

[52] L. R. e. M. H. David Wood, Marsha Zaidman, Linked Data. Structured Data on the Web, 2014.

[53] H. Patni, C. Henson, and A. Sheth, "Linked sensor data," in Collaborative Technologies and Systems (CTS), 2010 International Symposium on. IEEE, 2010, pp. 362–370.

[54] P. Barnaghi, M. Presser, and K. Moessner, "Publishing linked sensor data," in CEUR Workshop Proceedings: Proceedings of the 3rd International Workshop on Semantic Sensor Networks (SSN), Organised in conjunction with the International Semantic Web Conference, vol. 668, 2010.

[55] P.-Y. Vandenbussche, G. A. Atemezing, M. Poveda-Villalón, and B. Vatant, "Lov: a gateway to reusable semantic vocabularies on the web," Semantic Web Journal, 2015.

[56] A. Gyrard, C. Bonnet, and K. Boudaoud, "Helping IoT application developers with sensor-based linked open rules," in SSN 2014, 7th International Workshop on Semantic Sensor Networks in conjunction with the 13th International Semantic Web Conference (ISWC 2014), 19–23 October 2014, Riva Del Garda, Italy, 10 2014.

[57] A. Khandelwal, I. Jacobi, and L. Kagal, "Linked rules: principles for rule reuse on the web," in Web Reasoning and Rule Systems. Springer, 2011, pp. 108–123.

[58] S. Speiser and A. Harth, "Towards linked data services," in Proceedings of the 9th International Semantic Web Conference (ISWC). Citeseer, 2010.

[59] S. Speiser and A. Harth, "Taking the lids off data silos," in Proceedings of the 6th International Conference on Semantic Systems. ACM, 2010, p. 44.

[60] B. Norton and R. Krummenacher, "Consuming dynamic linked data." in COLD, 2010.

[61] Designing Cross-Domain Semantic Web of Things Applications Amelie Gyrard's PhD Thesis, Eurecom, Sophia Antipolis, 24 April 2015.

[62] Reusing and Unifying Background Knowledge for Internet of Things with LOV4IoT 4th International Conference on Future Internet of Things

and Cloud (FiCloud 2016), 22–24 August 2016, Vienna, Austria Amelie Gyrard, Ghislain Atemezing, Christian Bonnet, Karima Boudaoud and Martin Serrano.

[63] LOV4IoT: A second life for ontology-based domain knowledge to build Semantic Web of Things applications 4th International Conference on Future Internet of Things and Cloud (FiCloud 2016), 22–24 August 2016, Vienna, Austria Amelie Gyrard, Christian Bonnet, Karima Boudaoud and Martin Serrano.