Reusing and Unifying Background Knowledge for Internet of Things with LOV4IoT

Amelie Gyrard*, Ghislain Atemezing§, Christian Bonnet†, Karima Boudaoud‡, Martin Serrano*,
*Insight Center for Data Analytics, NUI Galway, Ireland, Email: {amelie.gyrard, martin.serrano}@insight-centre.org
†Eurecom, Sophia Antipolis, France, Email: {bonnet}@eurecom.fr
‡Laboratoire I3S-CNRS/UNSA, Sophia Antipolis, France, Email: karima@polytech.unice.fr
§Mondeca, Paris, France, Email: ghislain.atemezing@mondeca.com

***Abstract*—Ontology-based applications are becoming more and more popular and are usually domain-specific (e.g., eHealth or domotic). Designing ontologies and semantic-based applications manually is tedious and cumbersome for non-technical expert or semantic web beginners. Internet of Things (IoT) is a new field aiming to connect the physical world surrounded by devices such as sensors to the web to automatically interact with them and build innovative applications. The main challenges are to automate as much as possible the tasks of: (1) reusing the background knowledge previously designed by domain experts, (2) facilitating the tasks of IoT developers willing to integrate semantic web technologies into their applications, and (3) designing interoperable semantic-based IoT applications.**

Stemming from Linked Open Data and Linked Open Vocabularies, we designed Linked Open Vocabularies for Internet of Things (LOV4IoT), a catalogue of ontologies/datasets/rules relevant for IoT available online. LOV4IoT has been extended with more domains and ontology-based projects, a semantic-based dataset and a bot to enhance automation, and web services. Moreover, we demonstrate several use cases of the LOV4IoT dataset: (1) building Semantic Web of Things applications, (2) extracting frequent terms used in existing ontologies, and (3) stakeholders who can exploit, reuse and combine domain ontologies. Finally, we evaluated this dataset with users who exploit this dataset for their own purposes.

***Keywords*-Semantic Web Technologies; Linked Open Vocabularies; Internet of Things (IoT); Semantic Web of Things (SWoT); Domain ontologies; Ontology catalogue; Dataset catalogue; Rule catalogue; Interoperability.**

I. Introduction

Designing an ontology is becoming more and more popular. If the same ontology is not used in all systems, having an ontology will not solve the interoperability issues. The main challenging task it to reuse as much as possible existing ontologies in new semantics-based applications [8]. Moreover, usually ontologies are used for a specific application, the expertise already designed should be reused and combined to encourage interoperability among projects and applications. In this paper, we addressed the main research challenges:

- How to find and reuse the domain knowledge expertise to interpret sensor data? In theory, it could be done using semantic search engines and ontology/dataset catalogues. Due to existing limitations of such tools applied to the IoT domain, we built the Linked Open Vocabularies for Internet of Things (LOV4IoT)[1], a extension of Linked Open Vocabulary (LOV) [11].
- How to reuse, combine and unify domain knowledge (ontologies, datasets, and rules)? In theory, it could be done using ontology matching tools. However, finding the ontology matching tool fitting our needs among more than 100 tools is cumbersome and due to the existing limitations applied to the IoT domain, we redesign manually an interoperable domain knowledge and make it publicly available. To achieve this task, we use the idea of 'ontology network' [10] and 'modular ontologies' [1], to extract only one part of the ontology that we are interested in.
- Which semantic-based reasoning mechanism (rule-based inference, recommender system, machine learning) should we integrate to interpret sensor data? Within the IoT community and related applicative domains, most of the time logic-based and inference engine are used, more precisely the Semantic Web Rule Language (SWRL) [7].
- How to extract, reuse and combine the reasoning mechanisms found in existing domain knowledge? To address this, we build a dataset of interoperable rules to deduce meaningful information from IoT data.

The first contribution of this paper is the Linked Open Vocabularies for Internet of Things (LOV4IoT) which has been originally designed to assist developers in designing Semantic Web of Things applications to easily interpret Internet of Things (IoT) data. The main purpose of LOV4IoT is to: (1) reuse domain knowledge, (2) combine domain knowledge, and (3) extract rules to interpret IoT data. Lessons learned from this experience, we highlighted the need of an interoperable domain knowledge since we scrutinized interoperability issues.The previous version of LOV4IoT was a classification of the domain knowledge relevant for IoT represented as an HTML web page evaluating its reusability. In this paper, we enrich the LOV4IoT dataset as follows:

- It has been enriched with more projects and domains such as agriculture and smart energy. The dataset comprises almost 300 ontology-based projects and 19 domains.

[1] http://www.sensormeasurement.appspot.com/?p=ontologies

- It enables real-time statistics on the LOV4IoT dataset. It is now a semantic-based dataset and we can easily interact with this dataset through queries, etc.
- It is composed of the LOV4IoT bot to encourage domain experts to share their background knowledge.
- It can be reused by different stakeholders such as ontology matching experts, domain experts, semantic web experts, standardization experts and IoT experts. For instance, ontology matching tools should try their matching with this dataset. In the LOV4IoT dataset, ontologies are not structured in the same manner than usually.
- It classifies projects covering several domains such as smart home & energy or smart home & healthcare (Ambient Assisted Living). In the old version, it was difficult to classify such projects since they are involved in several domains.
- It provides web services which are used in the user interface to automatically compute the number of ontology by domains or according to their status (e.g., available online).

Moreover, the second main contribution of this paper is showing the way the domain knowledge has been extracted to build an interoperable knowledge required to design interoperable semantic-based IoT applications.

The paper is organized into sections as follows. Section II discusses related work. Section III highlights and explains the enriched LOV4IoT dataset. Section IV demonstrates the use of LOV4IoT to extract domain knowledge and make it interoperable and publicly available. Section V demonstrates use cases exploiting the dataset and main stakeholders who can benefit from this dataset. Section VI evaluates the dataset. Section VII highlights lessons learned from this work. Section VIII concludes this paper.

II. State of the Art

Developers could use semantic web tools indexing domain knowledge such as the Linked Open Vocabularies (LOV)[2] [11] catalogue for ontologies, the DataHub[3] project for datasets or semantic search engines such as Sindice, Watson[4] and Swoogle[5]. LOV does not reference ontologies which do not follow their best practices. Their ontologies are mostly designed by semantic web experts. READY4SmartCities[6] is a project to reduce energy consumption and CO2 emission in smart cities by using ontologies and linked data [9]. They build a dataset with 50 domain ontologies specific to smart cities and smart home. They do not cover main IoT domains such as healthcare, smart farm, etc. Further, they do not encourage domain experts to share online and improve their ontology according to the best practices. Table I compares the existing catalogues or tools referencing ontologies.

Tools	Feature	Pros	Cons
LOV	Ontology catalogue	- More than 469 ontologies referenced - Ontologies designed by semantic web experts - Bot to check best practices followed - APIs provided	- Not referenced if LOV recommendations are not followed - IoT domain limited
DataHub	Dataset catalogue	- 9,195 datasets - Various format accepted	- IoT domain limited - No quality checked - Manual referencing
READY4 SmartCities	Ontology & dataset catalogue	- More than 50 projects referenced	- Manual referencing
LOV4IoT (previous version)	Domain knowledge relevant for IoT	- More than 200 projects referenced - Ontologies designed by domain experts	- Manual referencing
Sindice, Watson, Swoogle	Semantic Search engines	- Automatic referencing	- IoT domain limited - Project not referenced if knowledge not available on the web.

Table I: Tools referencing domain knowledge

Most of the domain ontologies related to smart cities are not referenced on these tools since domain experts do not publish their ontologies online and do not follow semantic web guidelines. We scrutinize the following main challenges: (1) need of new tools to index domain knowledge expertise (ontologies and datasets) relevant for IoT, (2) catalogue of rules to interpret IoT data, (3) classification of sensors employed in IoT applications, (4) classification of technologies employed in IoT applications, (5) combining and reusing domain knowledge relevant for IoT, and (6) make the existing domain knowledge interoperable.

III. Linked Open Vocabularies for Internet of Things (LOV4IoT) dataset

In this section, we describe our first contribution: the Linked Open Vocabularies for Internet of Things (LOV4IoT) designed to enable reusing background knowledge and facilitating semantic-based IoT application development. We pursued a deeper analysis of domain knowledge involving sensors and semantic web technologies and came up with the following research questions: (1) Which sensors or actuators are employed? (2) What domains do sensors use? (3) Which ontologies exist that cover each domain? (4) What reasoning exit that cover each domain to interpret sensor data? (5) Is the ontology publicly accessible e.g., downloadable from a website? (6) Which technologies or tools are used to implement the ontology or rules? (7) Does the ontology follow the semantic web best practices? (8) Which projects could be reused and combined to other projects? (9) Which security mechanisms are used in the project?

LOV4IoT is the result of an analysis of the literature integrating sensors and ontologies into their project. LOV4IoT references almost 300 ontology-based works related to sensors in 19 domains: smart energy, activity recognition, weather, sensor networks, emotion, music, environment, fire,

[2]http://lov.okfn.org/dataset/lov/
[3]http://datahub.io/
[4]http://watson.kmi.open.ac.uk/WatsonWUI/
[5]http://swoogle.umbc.edu/
[6]http://www.ready4smartcities.eu/

health care, building automation, food, agriculture, tourism, security, transportation, smart city, IoT, Semantic Sensor Network (SSN), unit and other. More than 125 ontologies are now online and theoretically, could be easily reused. We discover, identify, study and reference these IoT projects since: (1) sensors and their measurements are described, (2) they can be used to design new cross-domain use cases (e.g., the naturopathy application to combine health, weather and smart kitchen), (3) the projects are based on ontologies, (4) the projects designed rule-based systems, (5) domain experts published their works in conferences, (6) they explained why they integrate semantics, (7) they described how they evaluate ontologies, (8) the ontology or dataset code could be used to implement new applications.

We analyze these works to reuse their ontologies and reasoning mechanisms. Most of the ontologies have been designed with semantic web standard languages such as RDF, RDFS and OWL. Moreover, frequently, the Semantic Web Rule Language (SWRL) has been used for the reasoning.

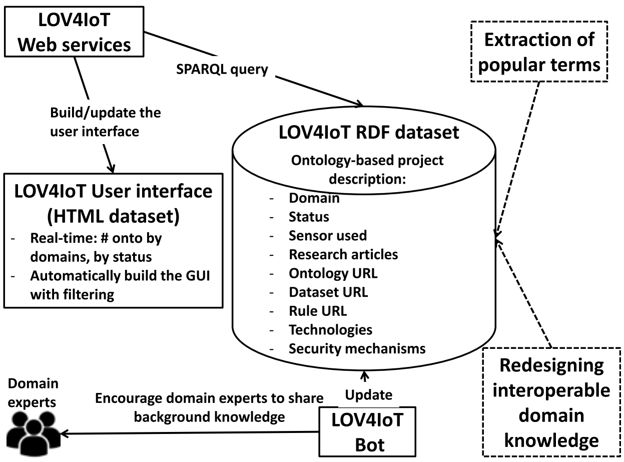

Figure 1: LOV4IoT overview

Figure 1 highlights the main components exploiting the LOV4IoT dataset that we describe in this section such as the HTML user interface, the bot, the web services. In section V, we also describe use cases with the functionality enabling the extraction of popular terms or re-designing an interoperable domain knowledge.

A. LOV4IoT, an extension of the LOV catalogue

LOV4IoT is an extension of the LOV catalogue [11], since the ontologies that we classified do not meet the requirements preconized by the LOV catalogue. The ontologies that we referenced in this dataset are not necessarily shared online, but we would like to exploit the knowledge expertise mentioned in the research articles. Requirements preconized by the LOV community such as ontology metadata or adding labels and comments to each concept and property are almost never respected. We contributed to the LOV community, to spread their best practices and encourage the 'sharing and reusing domain knowledge' approach. Unfortunately, we have seen that convincing authors to improve their ontologies is really a time-consuming task.

This limitation could be overcome by improving ontology editors to encourage people to add labels and comments. Recently, a beta version of ProtegeLOV[2][7] has been released, an extension of the popular ontology editor which suggests popular ontologies referenced in the LOV catalogue when you are designing a new concept or a new property.

The users can directly reuse the concept or integrate `owl:equivalentClass` or `owl:equivalentProperty` links. However, this plugin does not encourage users to add ontology metadata or labels and comments as preconized by the LOV community yet.

For these reasons, we build our own dataset, called Linked Open Vocabularies for Internet of Things (LOV4IoT) to reference and classify ontologies relevant for Internet of Things to: (1) interpret sensor measurements, and (2) combine domains. In our own dataset, we describe the ontology status according to the LOV criteria.

B. The LOV4IoT user interface

At the beginning of this work, we had few ontology-based projects referenced. We have chosen to classify them in a table and have indicated in each column: the authors, the date of publication, the related research articles, the sensors used, the technologies used, the rules employed and the security mechanisms. Each row in the table has a color to describe the status of the ontologies and rules: (1) lost or confidential, (2) we do not know if we can get the implementations, (3) the authors told us that they will share their knowledge expertise online, (4) the domain knowledge is shared online but do not meet he requirements preconized by the LOV catalogue, (5) the knowledge expertise is shared online and is even referenced on the LOV catalogue since they follow the required best practices. At the beginning, it was just a table in a word document. But, we thought that it can benefit to other people interesting in such ontologies, so we decided to share this classification. The word document has been converted to an HTML web page, by keeping the same idea with the table. It was easier and more flexible to add a new project or move a project to a new domain. Then the number grows more and more. So, we decided to split and classify them by domains.

The first column is dedicated to authors and the second to the publication date of the work. In the third column related publications are indicated and in the fourth column the ontology URL if it is provided is given. The fifth column indicates technologies used and the sixth column gives sensor used in the project. Finally, in the seventh column are indicated the rules designed in the projects (e.g., if foggy

[7]http://labs.mondeca.com/protolov/index.html

then safety devices are fog lamp, ESP and ABS). Further, each project is colored according to the ontology status such as the ontology is confidential or lost (in red), try to contact authors to get their ontologies or rules (in white), the authors will soon publish the ontology (in orange), the ontology is online (in yellow), online and referenced in LOV since semantic web best practices are followed (in green). Users such as developers, research engineers or domain experts can surf on this web page to search domain ontologies according to a specific domain.

Further, each project is colored according to the ontology status such as the ontology is confidential or lost (in red), try to contact authors to get their ontologies or rules (in white), the authors will soon publish the ontology (in orange), the ontology is online (in yellow), online and referenced in LOV since semantic web best practices are followed (in green).

C. LOV4IoT RDF dataset

Initially, the LOV4IoT RDF dataset has been designed to automatically compute the total number of ontology-based projects, the number of ontology per domains, and the number of ontologies according to their status (online, lost, publishing process online, referenced on LOV and contacting authors). It also enables to deal with projects covering several domains (e.g., smart home and weather). Moreover, the number of ontology-based projects attained 200 and the number of domains was growing.

The RDF LOV4IoT dataset[8] is available online and hosted on Google App Engine which ensures sustainability of this dataset. An extract of the LOV4IoT dataset in RDF/XML is depicted in Listing 1. Users such as developers, research engineers or domain experts can make statistics on this dataset or filter ontology-based projects by ontology status or by domains. The dataset also enables to automatically build a table in the HTML web page, to display a subset of the LOV4IoT dataset according to their needs. Machines can navigate on the RDF LOV4IoT dataset to easily retrieve the domain knowledge fitting their needs.

Creating SPARQL queries can be easily written to automatically count the total number of ontologies referenced within the RDF dataset.

We can even count the number of ontologies of each domain and according to the best practices status:

as depicted in Listing 2, SPARQL query 2.

- The total number of ontologies as as depicted in Listing 2, SPARQL query 2. The result provided by the execution of the query explains that 297 ontology-based IoT projects have been classified.
- The number of ontologies for each domain (Listing 2, SPARQL query 1). Ontology-based IoT projects have been classified by domains as follows: 46 for smart home, 8 for smart energy, 10 for activity recognition, 30 for tourism, 32 for transportation, 17 for agriculture, 16 for weather, 6 for smart cities, 21 for sensor networks, 55 for healthcare, 30 for food, 6 for affective sciences, 6 for music, 9 for environments, 7 for fire, 29 for security, and 5 other ontologies. Listing 2, SPARQL query 3 shows how to retrieve all ontologies URL from all projects in the IoT and sensor networks domains.
- The total number of ontology have been classified by ontology status as follows: 25 ontologies cannot be shared online (lost, confidential or not implemented), 121 ontologies have been referenced thanks to the research articles that we found but the ontology is not shared online, we are trying to convince authors to share the ontologies or rules on the web, 26 ontologies should be shared online soon according to the authors, 102 ontologies have been published online, most of them because we encouraged authors, 14 ontologies are online and referenced by LOV since the LOV best practices are adopted, and 9 ontologies were already referenced on LOV.

[8]http://www.sensormeasurement.appspot.com/dataset/lov4iot-dataset

D. LOV4IoT bot

We also have the LOV4IoT bot which encourages domain experts to share their ontologies, datasets or rules. The bot enables to send emails to domain experts to ask them to share their domain knowledge (ontologies, datasets or rules). This bot facilitates us the task when we are reading an article mentioning a new ontology, dataset or rule-based reasoning which has not been referenced yet on LOV4IoT, and could be useful to reference it and exploit it in further ontology-based IoT applications. This bot is a web service implemented in Java, using the JavaMail library, which has been used in our user interface.

E. LOV4IoT web services

LOV4IoT web services have been designed to automatically count the number of ontologies:

- `/lov4iot/totalOnto/` executes a SPARQL query to count the total number of ontology-based project referenced in the LOV4IoT RDF dataset.
- `/lov4iot/ontoStatus/{status}` executes a SPARQL query to count the different status of ontologies.
- `/lov4iot/nbOntoDomain/{domain}` executes a SPARQL query to count the different ontologies in each domain.
- `/lov4iot/sendEmail/{recipient, paper}` automatically sends an email to authors for a specific article.

For further detail, the LOV4IoT documentation is available online [9].

[9]http://sensormeasurement.appspot.com/documentation/LOV4IoTDocumentation.pdf

Listing 1: An extract of the LOV4IoT dataset in Turtle.

```
@prefix m3: <http://sensormeasurement.appspot.com/m3#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix lov4iot: <http://sensormeasurement.appspot.com/ont/m3/lov4iot#> .
<http://sensormeasurement.appspot.com/PaulStaroch>
 a <http://sensormeasurement.appspot.com/m3#M2MApplication> ;
 m3:hasContext m3:Weather, m3:BuildingAutomation ;
 rdfs:label "[Paul Staroch 2013]. See LOV4IoT for more details."@en ;
 rdfs:comment "Master’s Thesis: A weather ontology for predictive control in smart homes. 2013"@en ;
 m3:hasM2MDevice m3:Thermometer, m3:PrecipitationSensor, m3:HumiditySensor, m3:AtmosphericPressureSensor;
 m3:hasUrlOntology <http://paul.staroch.name/thesis/SmartHomeWeather.owl> ;
 m3:hasUrlRule <http://paul.staroch.name/thesis/SmartHomeWeather.owl> ;
 lov4iot:hasOntologyStatus lov4iot:OnlineLOV .
```

Listing 2: SPARQL queries to interact with the LOV4IoT RDF dataset

```

PREFIX m3: <http://sensormeasurement.appspot.com/m3#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX lov4iot: <http://sensormeasurement.appspot.com/ont/m3/lov4iot#>

#SPARQL QUERY 1: Get number of ontologies for each domain
select DISTINCT (count(?iotAppli) as ?nbOntoDomain) where{
   ?iotAppli rdf:type m3:iotApplication .
   ?iotAppli m3:hasContext ?domain .
}

#SPARQL QUERY 2: Get the total number of ontologies
select DISTINCT (count(?iotAppli) as ?ontologyTotal) where{
   ?iotAppli rdf:type m3:iotApplication .
   ?iotAppli lov4iot:hasOntologyStatus ?ontologyStatus .
}

#SPARQL QUERY 3: Retrieve all ontology URL from all projects in the IoT or SSN domains
select DISTINCT ?project ?ontology ?labelProject where{
   ?iotAppli m3:hasURLOntology ?ontology.
      {?iotAppli m3:hasContext m3:IoT}
   UNION
      {?iotAppli m3:hasContext m3:SensorNetworks}
}
```

F. Automatically Updating LOV4IoT

Updating the LOV4IoT dataset is simple, it is adding a new row in the HTML web page or a new instance in the RDF LOV4IoT dataset. If required, we could find additional background knowledge by connecting LOV4IoT to semantic search engines and ontology or dataset catalogue. At the beginning of this work, we started to use ontology catalogues such as Linked Open Vocabulary (LOV) [11] since it provides web services. Unfortunately, when we were experimenting this, we realized that most of the ontologies designed for IoT were not referenced on such tools yet.

IV. Extracting Domain Knowledge from LOV4IoT

LOV4IoT has been used to extract the domain knowledge by: (1) designing a dictionary to unify IoT data, (2) classifying popular applicative domains, and (3) designing a dataset of interoperable rules to deduce meaningful information from data. The extraction of the domain knowledge is depicted in Figure 2.

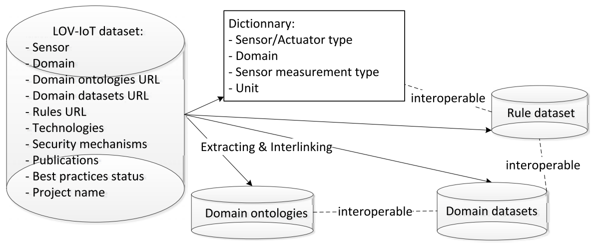

Figure 2: The LOV4IoT dataset used to build an interoperable domain knowledge

A. Extracting a dictionary to unify IoT data

Based on this work of classification and synthesization, a dictionary has been built to describe popular: (1) sensor or actuator type, (2) domain, (3) sensor measurement type, and (4) unit. For each sensor measurement, the corresponding unit is associated. The most popular term has been chosen since frequently synonyms describing a same sensor or sensor measurement are employed.

B. Extracting domains

We classified and referenced the most popular domains that we found in ontology-based projects more precisely: building automation, health, weather, agriculture, environment, emotion, transport, energy, tourism, location, city, tracking good (e.g., tracking food and tracking CD). In the building automation domain, the subclass activity recognition has been defined and in the environment domain, the subclass fire has been added. These domains are described in the dictionary and are used in: (1) the IoT application template dataset, (2) the LOV4IoT dataset to select or count the number of ontologies for each domain, (3) the drop down-list of the IoT application template generator when asking to choose a domain, (4) the semantic annotator to delete ambiguities and explicitly add the context to IoT measurements, (5) the web services, and (6) in the classification of IoT scenarios.

C. Extracting rules to interpret IoT measurements

We have analyzed that most of the ontology-based projects referenced in LOV4IoT employ inference engine, more precisely Semantic Web Rule Language (SWRL) [7] to infer new information from data produced by devices such as sensors.

We have classified and synthesized different languages that have been employed in the ontology-based projects. Heterogeneous languages are employed such as RIF, SWRL, `owl:Restriction` and SPIN. The SWRL language is frequently used by domain experts, since tools have been integrated in the popular Protege ontology editor tool. However, we analyzed the heterogeneity of SWRL rules. Indeed, SWRL rules are not interoperable because of the heterogeneity among software implementing SWRL syntax and reasoning engines. For instance, 6 Protege plugins have been designed (e.g., SWRL Tab, SWRL DL Safe Rule, SWRLJess Tab, SWRL-IQ, SQWRL and SWRLDroolsTab) to support SWRL within various reasoning engines. The rules can be described as `owl:Restriction` directly in the ontology. This is another kind of implementation of SWRL. Frequently, the rules that we are interested in are implemented in this way. SPIN is another language to describe rules, produced by authors using the TopBraid ontology editor. We referenced only 5 ontology-based projects mentioning this language.

Frequently, we found rules described as `owl:Restriction` directly in the ontology. For instance, a rule describes that if the precipitation measurement is equal to 0 millimeter per hour, then it does not rain. This rule has been previously described in a weather ontology. We expected that all rules are described in this way in all ontologies, but sometime the unit is not mentioned or the term mentioned to describe the sensor measurement type or the unit does not match to our dictionary, etc.

This heterogeneity hinders the automatic extraction of rules. For this reason, we manually redesign our own dataset of interoperable rules to infer high-level abstractions from sensor data. Such rules are based on the dictionary explained above. Since, most of the existing works design SWRL rules, we use the SWRL language to reuse and combine rules to enrich IoT data. When the rule dataset attained more than 100 rules, it has been split into sub-datasets to classify rules by domains: (1) health, (2) home, (3) weather, and (4) environment. This work is extensible with more rules and more domains. The rule dataset is exploited to build different semantic-based IoT applications.

V. USE CASES

The LOV4IoT dataset has been used in two use cases: the Machine-to-Machine Measurement (M3) framework to build interoperable Semantic Web of Things applications and the FIESTA-IoT EU project[10] to design an ontology by studying and aligning several IoT ontologies referenced on LOV4IoT such as IoT-Lite, M3, W3C SSN and QUDT. Moreover, we explain that different stakeholders can benefit from or exploit the LOV4IoT dataset such as domain experts, ontology matching tool experts, knowledge extraction experts, Semantic Web of Things developers and projects as depicted in Figure 3: **Machine-to-Machine Measurement (M3) framework** employs the LOV4IoT dataset to redesign interoperable domain ontologies, rules and datasets to assist IoT developers in designing semantic-based IoT applications without having to learn semantic web technologies thanks to the Machine-to-Machine Measurement (M3) framework [5]. **Domain experts** can use this dataset for their state of the art and to reuse existing ontologies or before designing their own ontologies, etc. For instance, a security expert used the LOV4IoT user interface to analyze existing ontologies in the security domain. **Knowledge extraction experts** can benefit from the LOV4IoT dataset since the domain knowledge expertise is referenced and classified. There is a need of

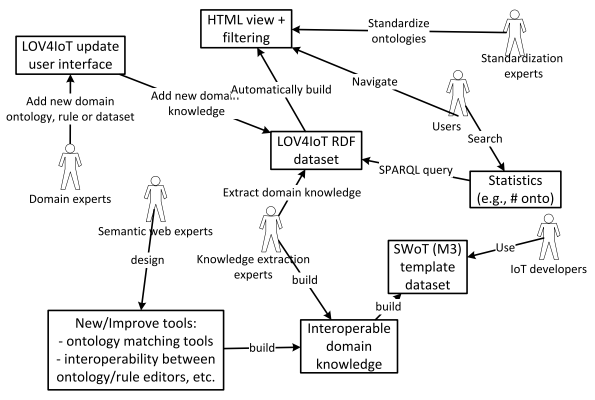

Figure 3: Use cases to exploit the LOV4IoT dataset

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

innovative tools to extract rules, etc. and redesign ontologies, datasets and rules in an unified way and having the same structure to facilitate interoperability in future architectures and systems. **Ontology matching tool experts** can reuse this dataset to later standardized the most popular and well-designed ontologies. **Ontology matching tool experts** benefit from LOV4IoT by analyzing interoperability issues explained in [3] and [4], exiting tools need to be improved to ease interoperability. Ontology editor tools such as Protege could preconize the re-use of existing ontologies based on the LOV and LOV4IoT dataset. When the user designed a new concept or property, some recommendations could be provided to reuse existing ontologies. This task is under development within ProtégéLOV[11]. Such extensions could be improved to recommend to integrate labels, comments, ontology metadata, etc. as recommended by LOV. Usually, ontology matching tools are evaluated with the Ontology Alignment Evaluation Initiative (OAEI)[12] benchmark. Current ontology matching tools are not adapted to ontologies referenced in the LOV4IoT dataset. A main challenge would be to have ontology matching tools adapted to both datasets (OAEI and LOV4IoT) meeting these main requirements: (1) heterogeneous languages, (2) syntactic heterogeneity, (3) conceptual heterogeneity, (4) terminological heterogeneity, and (5) semiotic heterogeneity. For instance, concepts or properties do not have labels or comments whereas most of the ontology matching tool algorithms are based use labels for the mapping. **IoT/SWoT developers and projects** can surf on the LOV4IoT web page to search domain ontologies according to a specific domain. For instance, the developer is looking for smart home ontologies, he goes on this section and finds more than 45 projects describing sensors and rules employed to build smart homes applications.

As explained above, the LOV4IoT dataset is relevant for various communities.

VI. EVALUATION

To evaluate LOV4IoT, an evaluation form[13] has been set up available on the LOV4IoT user interface. It has been filled by 35 volunteers who employed LOV4IoT, but additional volunteers can still provide their feedback. This form demonstrates that this synthesization and classification work of numerous ontology-based projects is useful for other developers, researchers and not only designed for the IoT research field. It helps users for their state of the art or finding and reusing the existing ontologies. Sometimes the results are not always 100% when the question was not mandatory or when we added later a new question to get more information. The LOV4IoT evaluation form contains the following questions and results.

- **Who are you?** Users are either: 43% semantic-based IoT developers, 40% IoT developers, 36% others, 7% ontology matching tool experts and 7% domain experts. It means that the dataset is mainly used by the IoT community. Among domain experts, we had feedback from security and IoT experts.
- **Domain ontologies that you are looking for?** 50% of users are interested in smart home ontologies, 35% in health ontologies, 27% in weather ontologies, 24% in transportation ontologies, 21% in security ontologies, 21% in some domains that we do not cover yet but not related to IoT, 18% in food-related ontologies, 12% in emotion, 9% in tourism, and 3% in agriculture. It means that users are interested in most of the domains that we cover.
- **How did you find this tool?** 42% found the LOV4IoT tool thanks to search engines, 36% thanks to emails that we sent to ask people to share their domain knowledge or to fill this form, 19% thanks to research articles, and 10% thanks to people who recommended this tool. Everybody can find and use this tool, not necessarily researchers.
- **Do you trust the results since we reference research articles?** 56% of users trust the LOV4IoT tool since we reference research articles, 41% are partially convinced. It means they consider this dataset as a reliable source.
- **In which information are you interested?** 67% of users are interested in ontology URL referenced, 64% of users are interested in research articles, 42% in technologies, 30% in rules and 28% in sensors used. The classification and description of each project is beneficial for our users.
- **Do you use this web page for your state-of-the art?** 34% of users answered yes frequently, 34% yes, and 31% no. Thanks to this work, users save time by doing the state of the art on our dataset.
- **In your further IoT application developments, do you think you will use again this web page?** 53% of users answered yes frequently, 41% yes, and 6% no. This result is really encouraging to maintain the dataset for domain and IoT experts.
- **In general, do you think this web page is useful**: 62% of users answered yes frequently, 35% yes, and 3% no. This result is really encouraging to maintain the dataset and add new functionalities.
- **Would you recommend this web page to other colleagues involved in ontology-based IoT development projects?** 82% of users answered yes frequently, 15% yes, and 3% no. This result is really encouraging to maintain the dataset and add new functionalities.

This evaluation shows that LOV4IoT is really relevant for the IoT community. The results are really encouraging to update the dataset with additional domains and ontologies.

[11] http://labs.mondeca.com/protolov/
[12] http://oaei.ontologymatching.org/
[13] https://goo.gl/mFUPVO

Moreover, according to Google Analytics, the LOV4IoT HTML web page has been seen more than 6 821 times (8 832 unique pages views) since August 2014. The average time spent on the web page is 3 minutes and half. It means that visitors returns several time to this dataset which demonstrates its usability.

VII. Lessons learned: Best Practices

Reusing and combining domain ontologies, datasets and rules as recommended by ontology methodologies [8] is really challenging. Due to heterogeneity, technical issues and limitations of ontology matching tools, reasoning engines and ontology or rule editors, the automatic extraction and the automatic linking of this domain knowledge is really difficult. To show the entire workflow to enrich sensor data with meaningful information, we manually re-design our own interoperable knowledge bases (rules, ontologies and datasets).

The automatic knowledge extraction is still considered as future work. Other research fields and tools could be used to overcome this challenge such as ontology design patterns, ontology methodologies, ontology merging & tools alignments, OWL 2 Rule Template or Natural Language Processing (NLP).

LOV4IoT demonstrates the need to spread semantic web best practices within the IoT community and related areas such as Web of Things, Smart Cities, Pervasive Computing, Context-Awareness, Ubiquitous Computing, etc. The set of semantic web best practices applied to IoT and related areas is summarized in [6].

VIII. Conclusion & Future work

LOV4IoT is a a tremendous work classifying and describing ontology-based projects into 19 domains and references more than 300 ontology-based projects. It has been extended with additional IoT applicative domains such as agriculture, smart energy, smart cities, IoT and Semantic Sensor Networks (SSN). LOV4IoT is not just an HTML web page anymore, but also a: (1) RDF dataset, (2) a bot to encourage domain experts to share their domain knowledge, and (3) web services to make statistics on it or to dynamically update the user interface.

The main novelty of the LOV4IoT dataset is encouraging the reuse of background knowledge to design interoperable semantic-based IoT applications. LOV4IoT has been evaluated with users and has demonstrated that numerous stakeholders can contribute or benefit from it such as domain experts, ontology matching tool designers, standardization experts and information retrieval experts.

Future work is to automatically extract the background knowledge to make it interoperable and to allow dynamic filtering though the user interface. LOV4IoT could also be integrated with semantic search engines and ontology/-dataset/rule catalogs to automatically update LOV4IoT.

Acknowledgment

The authors would like to thank the LOV team and Bernard Vatant for fruitful discussions regarding best practices and their valuable feedback. This work is partially supported by the Com4Innov Platform of Pole SCS[14] and FIESTA-IoT[15] from the European Horizon 2020 Programme with the Grant Agreement No. CNECT-ICT-643943.

References

[1] P. Doran. Ontology reuse via ontology modularisation. In *KnowledgeWeb PhD Symposium*, volume 2006. Citeseer, 2006.

[2] N. García-Santa, G. A. Atemezing, and B. Villazón-Terrazas. The protégélov plugin: Ontology access and reuse for everyone. In *The 12th Extented Semantic Web Conference (ESWC2015)*.

[3] A. Gyrard and C. Bonnet. Semantic Web best practices: Semantic Web Guidelines for domain knowledge interoperability to build the Semantic Web of Things, 04 2014.

[4] A. Gyrard, C. Bonnet, and K. Boudaoud. Domain knowledge Interoperability to build the semantic web of things, 06 2014.

[5] A. Gyrard, S. K. Datta, C. Bonnet, and K. Boudaoud. Cross-domain Internet of Things application development: M3 framework and evaluation. In *FICLOUD 2015, 3rd International Conference on Future Internet of Things and Cloud, August 24-26, 2015, Rome, Italy*, Rome, ITALY, 08 2015.

[6] A. Gyrard, M. Serrano, and G. Atemezing. Semantic web methodologies, best practices and ontology engineering applied to internet of things. In *WF-IOT 2015, World Forum on Internet of Things, 14-16 December 2015, Milan, Italy*, 2015.

[7] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean, et al. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member submission*, 21:79, 2004.

[8] N. F. Noy, D. L. McGuinness, et al. Ontology development 101: A guide to creating your first ontology, 2001.

[9] O. C. Raul Garcia-Castro, Asuncion Gomez-Perez. Ready4smartcities: Ict roadmap and data interoperability for energy systems in smart cities. In *2014 European Semantic Web Conference*, 2014.

[10] M. C. Suarez-Figueroa, A. Gomez-Perez, and M. Fernandez-Lopez. The neon methodology for ontology engineering. In *Ontology engineering in a networked world*, pages 9–34. Springer, 2012.

[11] 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.

[14]http://www.pole-scs.org/

[15]http://www.fiesta-iot.eu/