Enrich Machine-to-Machine Data with Semantic Web Technologies for Cross-Domain Applications

Amelie Gyrard, Christian Bonnet
Mobile Communication
Eurecom
Biot, France
Email: {gyrand.bonnet}@eurecom.fr

Karima Boudaoud
Rainbow team
Laboratoire I3S-CNRS/UNSA
Biot, France
Email: karima@polytech.unice.fr

Abstract—The Internet of Things, more specifically, the Machine-to-Machine (M2M) standard enables machines and devices such as sensors to communicate with each other without human intervention. The M2M devices provide a great deal of M2M data, mainly used for specific M2M applications such as weather forecasting, healthcare or building automation. Existing applications are domain-specific and use their own descriptions of devices and measurements. A major challenge is to combine M2M data provided by these heterogeneous domains and by different projects. It is really a difficult task to understand the meaning of the M2M data to later reason about them. We propose a semantic-based approach to automatically combine, enrich and reason about M2M data to provide promising cross-domain M2M applications. A proof-of-concept to validate our approach is published online (http://sensormeasurement.appspot.com/).

Keywords—Semantic Web of Things; Internet of Things; Machine-to-Machine (M2M); Semantic Web technologies; Domain Ontologies; Linked Open Rules; Linked Open Data; Cross-Domain Applications; Reasoning; Rules; SWRL; Naturopathy

I. INTRODUCTION

Interconnecting M2M data from heterogeneous domains such as healthcare, weather forecasting and building automation is a challenge since we loose the implicit information about the meaning of the M2M data. Indeed, we cannot apply the same reasoning on the human body temperature or the external temperature. The first temperature enables to deduce an eventual disease (e.g., fever), whereas the second one enables to deduce the weather conditions (e.g., cold). In the context of the Internet of Things, there is a need to explicitly describe the meaning of the M2M data provided by M2M devices. A second example deals with RFID tags embedded on goods (e.g., a bottle of milk) to obtain additional knowledge: the milk contains lactose, people can be allergic to lactose. By combining health, weather and smart kitchen M2M data, we can create promising cross-domain M2M applications such as a naturopathy application to suggest a recipe according to foods available in the kitchen, weather, user’s diseases, diets, allergies or emotional states. This scenario is depicted in Figure 1.

In this paper, we outline our contributions, more precisely, the semantic-based Machine-to-Machine Measurement approach (M3) to automatically combine, enrich and reason about M2M data to provide cross-domain M2M applications. The M3 approach includes the M3 ontology to define in a

Fig. 1. Combine cross-domain M2M data to suggest a recipe according to the weather and the user’s health

uniform way the various sensors, measurements, domains and units, the M3 hub to combine cross-domain knowledge and a new concept called Linked Open Rules to share and reuse in a uniform way semantic domain rules. Our M3 approach has been integrated in our semantic-based M2M architecture [1].

We introduce the necessity of a semantic-based approach to automatically enrich M2M data in section II. In section III, we present the state of the art. We explain the M3 approach: the M3 ontology, the M3 hub to build a cross-domain knowledge and the Linked Open Rules in section IV. Section V is dedicated to the prototype implementation. Finally, we conclude the paper in section VI.

II. THE HURDLE TO ENRICH M2M DATA

To combine cross-domain M2M data, it is essential to explicitly describe them. We propose to employ semantic web technologies to enrich M2M data in a unified way, as depicted in Figure 2. We explain in this section all hurdles encountered to semantically enrich cross-domain M2M data.

A. Getting M2M Data

Firstly, we have to deal with heterogeneous protocols to retrieve M2M data such as 6LowPAN, CoAP (Constrained Application Protocol), SenML [2] and SWE (Sensor Web Enablement) [3]. Existing works such as Semantic Sensor Web works (Sheth et al. [4]) are focused on the SWE (Sensor Web Enablement) which is a heavy and difficult to use protocol. In the context of our works, we have decided to use the
lightweight SenML protocol as it can be used easily and it is advocated by mostly the companies. According to the SenML draft, the protocol describes measurements: a measurement has a name, a value, a unit and a timestamp. The second assignment is that existing projects use their own notations to describe the name of the measurements, the domains and the sensor types. The term ‘t’, ‘temp’ or ‘temperature’ is used to describe the temperature measurement, ‘room 414’ or ‘office 414’ to express the domain, and ‘pluviometer’ or ‘rainfall sensor’ to represent the sensor. In the context of Internet of Things, more precisely, where M2M devices could communicate with each other without human intervention, there is a need to unify these terms. The human is able to understand those terms are for the same sensor, domain or measurement name but the M2M devices require an explicit and unified description of the measurements provided by semantic web technologies. For those reasons, we have designed the Machine to Machine Measurement (M3) ontology to classify and reference the various domains, measurements and sensor types to unify the description of the M2M data.

B. Combining Cross-Domain M2M Data

The second step is not an easy task since we want to combine the M2M data provided by heterogeneous domains and applications. Let is take the temperature measurement as depicted in Figure 2: the path A measured the body temperature whereas the path B measured the external temperature. Both are temperature measurements but the M2M device should not apply the same reasoning. In the first case, the M2M device may deduce that the person gets the flu, in the latter case, the M2M device concludes that the temperature is hot.

C. Reasoning on Data to Deduce a New Information

A significant interest of semantically annotating M2M data is to reason about them to provide cross-domain M2M applications. The main goal of the step 3, depicted in Figure 2, is to infer a new information by applying semantic rules. For example, if the body temperature is greater than 38°C and the person is located in the bedroom, she/he is probably sick. Currently, existing domain applications manually index all these semantic rules which are not published online. We propose a new concept called Linked Open Rules to reference, share and reuse semantic rules since they are already implemented by domain experts.

D. The Meaning of the New Information is Described in Domain Ontologies

In the previous step, the M2M device deduces the new concept ‘flu’ or ‘hot’ from the sensor value 38°C. Nevertheless, the meaning of this information is not interpreted yet, while it is already defined in domain ontologies. In the example, the ‘flu’ concept is referred to disease in health ontologies, and the ‘hot’ concept is related to season in weather ontologies. We propose that our M3 ontology acts as a hub to fuse cross-domain knowledge (ontologies and datasets).

E. Semantic Guidelines to Share the Domain Knowledge

Step 5 intends to find the most relevant datasets corresponding to the previous ontologies to get additional knowledge. Firstly, numerous interesting domain ontologies described in research articles are not published online. Secondly, if they are published online, they do not respect semantic web best practices. For example, these ontologies are not indexed by semantic web tools which reference ontologies such as the Linked Open Vocabularies catalogue (LOV) or the Watson semantic web search engine. There is also the useful Oops tool to detect common errors when creating ontologies. We encounter the same problems with domain datasets, the DataHub project, the Linked Open Data and the semantic search engines such as Sindice, Swoogle, Falcons that reference numerous datasets. Thirdly, we manually find interesting domain ontologies which are not designed considering the existing ones. For example, the smart home ontology for elderly people has the concept Person and the related properties hasName/hasFirstName to represent the patient, this concept is already defined in health ontologies. All of these common concepts should be linked to those already defined in existing ontologies. We encounter the same issue for datasets. Ontologies and datasets having common concepts or instances should be linked with the keyword owl:equivalentClass or owl:sameAs.

It is necessary to promote semantic web best practices and tools to domain experts to reuse their knowledge in numerous domains such as food, healthcare, building automation, emotion, movie, earthquake, tourism, security, agriculture and intelligent transport system.

F. Exploiting the Complementary of the Knowledge to Build a Cross-Domain Knowledge

Existing ontologies or datasets are independent, but numerous domain knowledge could be combined to exploit the complementary between the existing domains. A first example is that the Person concept is described in two ontologies: health and FOAF. In the health ontology it is described as patient, whereas in the second one it is described as a digital identity. In both ontologies, Person represents the same physical person through different point of views. Another example is that the Cabbage concept provided by the Smart Products

1 http://lov.okfn.org/dataset/lov/
2 http://oeg-lia3.dia.fi.upm.es/oops/index-content.jsp
3 http://datahub.io/
4 http://linkeddata.org/
5 http://xmlns.com/foaf/spec/
project6 is used to describe food related to recipes and the Cabbage concept is also described in an agricultural ontology. Exploiting the complementary of the domain knowledge is not an easy task since they are not described in the same manner. Firstly, the orange concept is described as a concept in an ontology and as an instance in a dataset. Secondly, some concepts or instances miss the common property rdfs:label. Neither the LogMap tool7 to link ontologies with each other using the owl:equivalentClass property, nor the Silk tool8 to link datasets using the owl:sameAs property can be used. We propose the M3 hub to exploit the complementary of the knowledge by linking cross-domain ontologies and datasets.

G. Summary

It is difficult to reuse the semantic-based domain knowledge since domain experts do not follow semantic web guidelines and do not reference their works on semantic web tools. How can we exploit on one hand the domain knowledge and on another hand the semantic web tools to automatically enrich the M2M data? We propose in section IV a semantic-based approach to automatically combine, enrich and reason about M2M data to create cross-domain applications. More precisely, we describe the M3 ontology, the M3 hub and the Linked Open Rules. One of our cross-domain scenarios is a naturopathy application to combine meteorology, home automation, affective science and healthcare domains.

III. RELATED WORKS

We present in this section the existing domain-specific M2M applications and semantic sensor networks.

A. Machine-to-Machine applications

We have been inspired by the ETSI M2M [6] architecture to build the semantic-based M2M architecture to semantically annotate the M2M data. Existing M2M applications are domain-specific such as home monitoring [7] [8], healthcare [9], intelligent transport system [10] or smart supermarket-fridge [11].

Existing M2M works do not provide semantic-based applications and are domain-specific.

B. Semantic Sensor Networks

Sheth et al. [4] propose the new concept of ‘Semantic Sensor Web’ to semantically annotate sensors and their data. The SemSOS [12], the Sense2Web platform [13] [14] [15] and the Sensor4grid4env [16] (Semantic Sensor Grids for Environmental Applications) semantically annotate sensor streams and link them to the Linked Open Data, mostly in the environmental domain (e.g., weather forecasting, fire prevention and flood control). The Spitfire [17] project coined the term ‘Semantic Web of things’ and focus on building automation domain.

Firstly, these related works focus on semantic annotation for the interoperability of sensors rather than the sensor data itself. Secondly, all of these works get sensor data through the SWE protocol, whereas we use the lightweight protocol SenML advocated by companies. Thirdly, these projects do not reuse knowledge provided by domain experts (affective science, intelligent transport system, agricultural, healthcare, tourism). They only reuse ontologies and datasets designed by semantic web experts. Fourthly, these projects are domain-specific and do not build cross-domain knowledge. Finally, they do not propose a semantic-based reasoning such as the new concept Linked Open Rules. To sum up, these existing works do not provide a semantic-based approach to combine, enrich and reason about the M2M data to design cross-domain M2M applications.

IV. THE PROPOSED M3 APPROACH TO ENRICH M2M DATA

To address the questions presented in section II, we propose the M3 (Machine-to-Machine Measurement) approach, as depicted in Figure 3, a novel approach to combine, enrich and reason about M2M data to build cross-domain M2M applications.

A. The M3 Ontology

In Figure 3, the first step retrieves heterogeneous M2M data according to the SenML protocol. The second step annotates M2M data with semantic web standards such as Resource Description Framework (RDF), RDF Schema (RDFS) and Ontology Web Language (OWL). RDF is based on triplets. A triplet is like a sentence (subject, verb and complement) called ‘subject, predicate, object’. For example, in the triplet ‘the thermometer is a sensor’, ‘thermometer’ is the subject, ’is a’ the predicate which means the property rdfs:type, and ‘sensor’ the object. RDF is insufficient to describe things, the RDFS standard adds additional information such as the notion of hierarchy (e.g., body temperature rdfs:subClassOf temperature). Finally, OWL is used to design your own ontology by creating new concepts and the relationships between them. For example, the concept Sensor, Measurement and the property produces have been created to explain that a sensor produces measurements.

We design the M3 ontology9, an extension of the Semantic Sensor Networks (SSN) ontology [18] to describe M2M data

---

6http://projects.kmi.open.ac.uk/smartproducts/
7http://csu6325.cs.ox.ac.uk/
8http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/
9www.sensormeasurement.appspot.com/m3#
in a uniform way. The M3 ontology: (1) defines main components in the M2M architecture, (2) semantically annotates M2M data, (3) classifies M2M devices, their data and the domains and (4) links the M3 concepts to domain ontologies to obtain additional information. We reused the SSN ontology since we have common concepts such as Sensor. We explain that an M2MDevice can be Actuator, RFIDTag, Controller, Transducer or Sensor. The M2M device observes a Domain and produces Measurements. A measurement has a name, a value, a unit, a type and a timestamp.

The M3 ontology describes more than 30 sensors (e.g., Thermometer, Light) and various Domains such as Health, Weather, Transport, Tourism, Building. We classify them in a hierarchy, for example, the pressure sensor can be a blood pressure sensor or an atmospheric pressure sensor and the environment domain can be weather, pollution or smart agriculture. We also add restrictions between the sensor, the produced measurement and their units. For example, the thermometer is a sensor which produces temperature measurements and the temperature unit is degree Celsius, Fahrenheit or Kelvin. We classify a great deal of sensor measurements, e.g., Temperature, Humidity, BloodGlucoseLevel, Illuminance. RFID tags are already embedded on CD, Clothes and DVD. In a foreseeable future, RFID tags will be embedded on numerous products such as Book and Food. So, we add numerous RFIDMeasurementType in the M3 ontology.

B. M3: a Hub for Cross-Domain Knowledge

In the previous section, we have semantically annotated the M2M measurements in an unified way using the proposed M3 ontology. To obtain additional information, we fuse the M3’s concepts to those found in existing domain ontologies designed by domain experts. We found sensor, weather, earthquake, health, smart building, emotion, tourism, transport, agricultural and food ontologies. Domain ontologies have been designed for specific applications without being linked to the previous ones. For example, ontologies in the same domain (e.g., healthcare) are not linked with each other. In complementary domains, ontologies are not linked with each other (e.g., link the food ontologies with agricultural ontologies).

For those reasons, we design in step 4 M3 as a hub for cross domain ontologies and datasets. We generate a new knowledge by interconnecting the new types inferred with those from domain ontologies. For example, the concept Flu or BodyTemperature is fused to the one proposed in health ontologies via the property owl:EquivalentClass. To update and automatize the M3 hub, we propose to use the Linked Open Vocabularies project to find and integrate a new well-designed domain ontology.

A similar approach is used to enrich semantic M2M data with the Linked Open Data in step 5. We use the property owl:equivalentClass to link instances with each other instead of the property owl:equivalentClass to link concepts with each other. The M3 approach integrates the dataset corresponding to the domain-specific ontology. We found interesting datasets related to disease, emotion, recipe, activity, weather, etc. Unfortunately these domain-specific datasets are not linked with each other and do not follow semantic web best practices. So, we build manually cross-domain datasets in step 6. For example, the naturopathy dataset is a cross-domain dataset to combine disease, food, recipe, emotion, color, season and nutrient datasets. To update and automatize the M3 Hub, we propose to use the DataHub project which references datasets to find and integrate a new domain dataset. Using the Silk tool, we can automatically add the property owl:sameAs between two datasets, and the LogMap tool automatically add the property owl:equivalentClass when the datasets or the ontologies are well-written and follow the semantic web best practices.

C. The Linked Open Rules

Step 3 is dedicated to the semantic-based rules to enrich the M2M data with a new information. The result of the rule is a new concept defined in the M3 ontology and linked to domain ontologies as explained in the previous section. An example is that the M3 approach deduces from the health domain and the temperature measurement that the M2M data corresponds to a BodyTemperature. Another rule deduces that if the BodyTemperature is higher than 38 °C and the person is located in the bedroom then it corresponds to a Flu. The Flu concept is described as a Disease in health ontologies.

In the step 3 and 7, we propose a new concept that we call Linked Open Rules as a cross-domain reasoning to share and reuse semantic rules in a uniform way. Semantic Web Rule Language (SWRL) is the common language used to define domain rules in various domains such as weather, healthcare, etc. SPARQL, which is the W3C standard to query semantic data, is a SQL-like language. An inference engine reasons on the properties owl:equivalentClass, rdfs:subClassOf, etc. As explained previously, the M3 approach deduces the concept Disease, which is described in the naturopathy ontology and dataset. For example, home remedy such as lemon and honey are recommended for the Flu. We create SPARQL queries in the step 7 and use the inference engine to provide cross-domain M2M applications such as the naturopathy application which suggests a recipe according to weather, user’s diseases and diets. Naturopathy is a cross-domain application since it combines meteorology, healthcare, affective science and smart kitchen domains.

D. Semantic Web Guidelines Summing-Up

Since domain experts are unaware of semantic web best practices, we propose to share lessons learned to build well-designed ontologies and datasets that we acquired with the help of semantic web experts. We design a web page\footnote{http://www.sensormeasurement.appspot.com/?p=bestPractice} to summarize semantic web guidelines and reference useful semantic tools. We also sent more than 80 emails to ontology-based research projects in disparate domains (affective science, intelligent transport system, weather forecasting, building automation, healthcare, security, tourism, agriculture, earthquake) to advise them to publish their ontologies, datasets and rules online. One third of the domain experts answered us and were thankful. Some of them are publishing their work online according to the semantic guidelines and are using the referenced semantic web tools. In the next paragraph, we propose
some lessons learned, presented on our website to build well-designed semantic-based applications.

1) Reference Domain Knowledge: To ease the development of future applications, domain experts should share their ontologies, datasets and rules on semantic web tools.

- Reference domain ontologies on the LOV catalogue\(^{11}\) [19] and the semantic search engines such as Watson and Swoogle.
- Reference domain datasets on the DataHub project\(^{12}\) and on semantic search engines such as Sindice\(^{13}\).
- Reference domain rules on the Linked Open Rules\(^{14}\) which is still a work in progress.

2) Semantic Guidelines:

- To have an ontology referenced on LOV:
  - Add the metadata descriptions proposed by LOV [19].
  - Add the properties rdfs:label and rdfs:comment.
  - Add the property owl:equivalentClass with the class already described and referenced on LOV.
- Follow the semantic web best practices to design an ontology [20] and use the OOPS project\(^{15}\) to detect common ontology pitfalls.
- Use the Linked Data principles\(^{16}\) to create a well-designed RDF dataset.

E. Limitations of the M3 approach

Some part of this M3 approach are already implemented and automatized: convert senML data into RDF data (1) using the M3 ontology (2). We intent to integrate semantic tools which are still works in progress to automatize the M3 approach: the Linked Open Vocabularies and the Linked Open Data (5). To automatize M3 is not a easy task since domain experts do not reference their ontologies and datasets on these tools. Finally, other parts of the M3 approach are research challenges and work in progress: the Linked Open Rules (3) (7), the hub for cross-domain ontologies (4) and datasets (6). This approach is the first approach to automatically combine, enrich and reason about M2M data to build cross-domain applications.

Since the M3 approach is linked to numerous domain ontologies and datasets, this approach is not entirely scalable and requires a high time-consuming process. To deal with this issue, we are introducing the notion of ‘ontology clustering’ to only use domains that you are interested in. To implement this process, a solution could be to enable/disable the matching (owl:sameAs, owl:EquivalentClass).

\(^{11}\)http://lov.okfn.org/dataset/lov/
\(^{12}\)http://datahub.io/fr/
\(^{13}\)http://sindice.com/
\(^{14}\)http://www.sensormeasurement.appspot.com/?p=rule
\(^{15}\)http://oeg-lia3.dia.fi.upm.es/webOOPS/index-content.jsp
\(^{16}\)http://linkeddata.org/

V. Prototype Implementation

To demonstrate the feasibility of the M3 approach, we have developed a prototype published online\(^{17}\) and we have implemented some cross-domain M2M applications (e.g., naturopathy, tourism).

A. Combine Weather, Healthcare, Emotion and Smart Home Domains

One of our scenarios of cross-domain M2M applications is a naturopathy application to suggest a recipe according to weather, user’s diseases and diets. This application combines cross-domain ontologies and datasets: weather, season, disease, diet, emotion, recipe, color, food and their nutrients. The naturopathy application has been split into three scenarios. A first scenario, depicted in Figure 4 simulates the 38°C temperature measurement provided by the health domain according to the SenML protocol. The M3 approach converts the SenML data into semantic data using the M3 ontology. The M3 rules referenced in the Linked Open Rules deduce that the M2M data is a body temperature and that the person has the flu. The flu is referred as a disease in the naturopathy and health ontologies which are referenced in the M3 ontology since M3 is a hub to fuse cross-domain ontologies. Once we find the naturopathy ontology, the M3 approach loads the corresponding naturopathy dataset which references food, nutrient, disease, season, color, etc. to propose a home remedy (honey, lemon, thyme, etc.). In the second scenario, the temperature is provided by the weather domain. The M3 approach infers rules to detect that the weather temperature is hot as in summer. The M3 approach, through the inference engine and SPARQL queries, suggests ingredients according to the season. A third scenario simulates food measurements: the M3 approach links food measurements to domain-specific ontologies, such as those provided by the smart products project and knowledge bases (recipes). This scenario, depicted in Figure 5, suggests a recipe according to food available in the smart kitchen.

B. Combine Weather and Tourism Domains

To validate that the M3 approach is generic, we propose two other cross-domain M2M applications. A first scenario deals with weather measurements such as temperature, precipitation, wind speed and luminosity. The M3 approach deduces...
Fig. 5. Suggest a recipe according to food available in the kitchen

Fig. 6. Suggest activities according to the weather

the actual weather (rainy, sunny or windy, etc.) and suggests an activity. As depicted in Figure 6, activities proposed are sunbathing, beach volley and fishing when the luminosity is high. If the weather is windy, possible activities proposed are windsurfing. A second scenario deals with location measurements: longitude and latitude. The M3 approach semantically annotates these measurements, and links them to the restaurant, geonames\textsuperscript{18} ontologies and datasets to suggest a restaurant around a specific location. This kind of restaurant suggestion already exists, but existing applications are not based on semantic web technologies. These scenarios show that the M3 approach is able to deal with diverse kinds of measurements and domains.

VI. CONCLUSION AND FUTURE WORKS

We proposed the M3 approach to combine cross-domain M2M data and enrich them with semantic web technologies. The M3 approach reasons about the semantic M2M data to suggest a restaurant according to the user’s activity and the actual weather (rainy, sunny or windy, etc.) and suggests an activity. As depicted in Figure 6, activities proposed are sunbathing, beach volley and fishing when the luminosity is high. If the weather is windy, possible activities proposed are windsurfing. A second scenario deals with location measurements: longitude and latitude. The M3 approach semantically annotates these measurements, and links them to the restaurant, geonames\textsuperscript{18} ontologies and datasets to suggest a restaurant around a specific location. This kind of restaurant suggestion already exists, but existing applications are not based on semantic web technologies. These scenarios show that the M3 approach is able to deal with diverse kinds of measurements and domains.

ACKNOWLEDGMENT

The authors would like to thank the Eurecom’s semantic web team (Ghislain Atemezing and Raphael Troncy) for fruitful discussions or help for the semantic web part, David Ramamonjisoa and colleagues for the reviews. This work is supported by the Com4Innov Platform of Pole SCS\textsuperscript{19}.

\textsuperscript{18}http://www.geonames.org/
\textsuperscript{19}http://www.pole-scs.org/

REFERENCES