Semantic Registries for Heterogeneous Sensor Networks

Bridging the semantic gap for collaborative crises management

Fernando Chaves, Jürgen Moßgraber, Manfred Schenk, Ulrich Bügel
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB
Karlsruhe, Germany
e-mail: {ferando.chaves-salamanca, juergen.mossgraber, manfred.schenk, ulrich.buegel}@iosb.fraunhofer.de

Abstract—The management of heterogeneous sensor networks is a complex task. Moreover, such networks often form an isolated system that needs to be connected to other network systems. Therefore, a system-of-systems approach is required. A sensor registry is the central part of a distributed sensor network. In this paper, we describe how to enhance sensor registries with semantic technologies in order to overcome some of their limitations. The Sensor Web Enablement standards of the Open Geospatial Consortium constitute a consistent and flexible framework in order to assure syntactical interoperability between applications, which make abundant use of data from sensor networks operated by heterogeneous stakeholders. The data model used throughout these standards is the Observation & Measurement model. The conversion into an ontology provides an easier handling of the model. In this paper, we propose the use of such an ontology as the basis of a Semantic Registry. Besides providing the information about the available sensor infrastructure and their semantically rich annotations, it can easily be extended to handle for example information about available services for data analysis or simulations. Furthermore, the resilience of the SRs is a crucial point in the system architecture to prevent it becoming a single point of failure.

Keywords—semantic registry; semantic sensor network (SSN) ontology; service orchestration; service choreography; ontology management; system-of-systems (SoS)

I. INTRODUCTION

In the past years, sensor network and sensor web applications based on the Sensor Web Enablement (SWE) standards of the Open Geospatial Consortium (OGC) have increased both in number and functional capabilities, e.g. for enabling decision support services for crisis management scenarios. These standards constitute a consistent and flexible framework for assuring syntactical interoperability between applications. However, both sensor suppliers and stakeholders of operational systems are still reluctant to adopt these standards in their products and systems shying the unavoidable a-priori adoption and adaptation effort and the uncertain benefits, e.g. for the collaborative management of evolving crises. Organizational, legal and political constraints tend to restrict free interchange of sensor metadata and data between different stakeholders. Therefore, few systems really make extensive use of sensor data from distributed networks operated by heterogeneous stakeholders. The components of a system are quite often optimized for specific purposes and do not take into account interoperability issues.

The EU Project TRIDEC – Collaborative, Complex and Critical Decision-Support in Evolving Crises [1] focuses on new approaches and technologies for intelligent geo-information management in complex and critical decision-making processes in Earth sciences. The key objective in TRIDEC is to design and implement a collaboration infrastructure of interoperable services, which must be able to manage intelligently information and data increasing dynamically in terms of both size and dimension. TRIDEC demonstrates the research results in two scenarios involving intelligent management of large volumes of data for critical decision-support. The first scenario Natural Crisis Management (NCM) concerns a large group of experts working collaboratively in crisis centres and government agencies using sensor networks. Their goal is to make critical decisions and to save lives as well as to protect infrastructure and industrial facilities in evolving tsunami crises. The second scenario Industrial Subsurface Development (ISD) concerns a large group of consulting engineers and financial analysts from energy companies working collaboratively in sub-surface drilling operations. Their common objective is to monitor the operations in real-time using sensor networks, to optimize the drilling processes and to detect upcoming critical situations in drilling operations as early as possible. Both scenarios share the requirement to manage a large, geodistributed and heterogeneous sensor network that changes dynamically. Typically, such a network consists of many sub-networks, the majority of which is vendor specific; these sub-networks operate independently and have to be seen as “black boxes” from a network designer point of view. As a response to this complexity, TRIDEC applies a system-of-systems (SoS) architecture.

In this paper, we describe the pragmatic approach followed in TRIDEC for mapping de-facto or established industrial metadata standards from TRIDEC’s application domains to elements of the semantic sensor network (SSN) ontology [9]. The main aim is to facilitate the combination of SWE and semantic web standards by crisis management organizations, drilling companies and sensor vendors in their operational components, systems and applications. We also discuss how interoperability between legacy applications can
be enhanced by means of Semantic Registries as a central component.

After presenting related work in Section II we outline requirements to overcome the obstacles named by industry representatives hindering the adoption of semantic technologies for decision support in general and, specifically, in the field of crisis management. In Section IV, we introduce the concept of a decision support ontology, which forms the basis of the semantic registry (SR). Section V presents the specific software implementation of the TRIDEC SR followed by a usage example from the TRIDEC drilling scenario. In the final section, we discuss the pros and cons of the architecture and implementation and conclude with some final remarks on future work.

II. RELATED WORK

The Universal Description Discovery and Integration (UDDI) specification, which was proposed as a service registry for Service-oriented architectures. It failed the adoption of big industrial companies and needs to be classified as obsolete. Its facilities for service discovery are rather limited due to the lack of machine-understandable semantics in the technical specifications and classification schemes.

The FUSION SR developed in the context of the EU-funded project FUSION [2] aimed at overcoming these limitations. It created semantically annotated service interface descriptions and used OWL-DL for modelling service capabilities and for performing matchmaking via description logic reasoning [3].

The publication of the final report of the W3C Semantic Sensor Network Incubator Group [4] made it feasible to follow a similar approach for integrating services and products in the sensor network domain, which suffers from similar shortcomings. Although substantial work had already been carried out towards providing an ontological representation of the Observations and Measurements (O&M) model [5], the report was the first to provide an extensive survey of existing sensor network applications and a consistent framework that offers a unified view on them. The report followed two main objectives:

1) to develop an ontology to describe sensors and sensor networks for use in sensor network and sensor web applications, and
2) to study and recommend methods for using the ontology to semantically enable applications developed according to available standards such as SWE.

Another approach which also uses the term “Semantic Registry” has been developed by the EU-funded project SemSorGrid4Env [7]. As part of this project, some of the fundamental specifications from the semantic web area have been extended to support spatial and temporal aspects. In contrast to this approach, we confined ourselves to the use of existing standards without modifications in order to encompass a larger set of existing systems.

III. USER REQUIREMENTS ON INFORMATION MANAGEMENT FOR DECISION SUPPORT

The requirements of the SR cannot be decoupled from the broader requirements of their usage in a decision support system based on sensor data. A recent survey on the use of semantic Web technologies for Decision Support [8] analysed the landscape of semantic technologies in Decision Support Systems (DSS). The survey compiled a number of previous reports and additionally conducted a small set of interviews in Sweden in order to capture a snapshot of viewpoints from non-university research and industry. The survey concludes that DSS is the class of applications that has the largest gap between existing solutions and perceived needs. The main issues for DSS today are – in this order: (i) interoperability, (ii) searching and linking of information, (iii) collaboration support.

These results correlate with our own experience gained in using ontologies in a Web portal for the remote monitoring of nuclear power plants in south-west Germany [12]. In order to group and classify these requirements, which motivated design and development of the TRIDEC SR, we refer to the system feature categories introduced for conducting the interviews in Sweden, and cite or summarize prominent results of the survey from our specific point-of-view:

A. Information Integration

Information Integration addresses the need for integrating sensor information with reports from human actors, as well as external information from the web. Although this need clearly exists, most often static connections between systems are set up including tailored transformation software that must be maintained as schemas or technologies of data sources change.

B. Information Filtering and Selection

In most systems there are several types of users with different roles and motivations for using the DSS, which results in different views. Today, these views are most often generated by a fixed set of filtering rules that are decided on at design time.

C. Information Extension, Exploration and Explanation

Most DSS today seem to lack an “open world view” and are rather static, as the sources are determined at design time and data is usually not linked or traceable. Explanations of derived information, and drill-down in terms of exploring the underlying data or its sources, seem to be uncommon.

D. Information Tracking and Post-Event Analysis

Tracking of large data flows is concerned with historical data, and related to issues such as data provenance, post-analysis of how a situation has unfolded, security and individual responsibility issues.

E. Models and Model Evolution and Decision Sharing

The lack of flexibility in models are a severe problem, whereby models also need to be dynamic and change along
with the evolving data, schemas and technologies in general. Similarly, the need for storing of meta-level information, i.e., information about decisions themselves, arises in several example scenarios.

These requirements needed to be addressed by the architecture and implementation of TRIDEC’s SR to solve the most acute information management problems mentioned above.

IV. THE DECISION SUPPORT ONTOLOGY (DSO)

The DSO forms the schema for the knowledge base of the SR. The main challenge in its design is to adequately adapt the concepts to the real world objects (e.g., sensors, data streams) and operational procedures, which govern the management of a crisis. As a basis, we use the Semantic Sensor Network Ontology (SSNO) that already describes sensors and sensor networks. It is aligned with the DOLCE+DnS Ultralight (DUL) in order to assist its use in conjunction with ontologies or linked data resources developed elsewhere, e.g., the Measurement Units Ontology (MUO). The design of the TRIDEC DSO is based on a top down approach by re-using and extending these domain independent ontological patterns by means of thematic models derived from use-cases found in the TRIDEC.

Figure 1 illustrates an excerpt of the DSO that illustrates these principles. It shows the relationships of the SSNO concept “Property”. Properties are qualities that can be observed by a certain kind of sensor; they inhere in Features of Interest, which are entities in the real world that are the target of sensing. A Property has relationships to classes defined by the upper ontologies (e.g. to SSNO concept “Observation”) and to subclasses which have been defined for the TRIDEC domains (“PropertyISD”, “PropertyNCM”). In the ISD domain, properties typically have different names depending on the users’ roles and views. Furthermore, existing standards need to be respected and integrated. Such a standard is the Wellsite Information Transfer Specification (WITS) for the transfer of drilling rig-related data. Layer 0 is an ASCII-based transfer specification. As an example the depth of the drilling bit can be identified by its WITS0 code for the corresponding data channel (“C0108A”) but can also be identified by additional long (“MDBit”) or even language-specific names (“BitDep”). Such equivalencies are handled in the DSO by means of OWL equivalent classes.

The current DSO is agreed as a common vocabulary among the domain experts (speaking different languages, have distinct responsibilities and different roles) and IT experts (lacking the necessary domain knowledge).

A DSO development cycle starts with competency questions to domain experts in order to achieve an informal specification. Formal specification and consistency checking is done with the Protégé editor. The resulting OWL file is then installed in the SR; which then provides automatic generation of input forms, an easy to use mechanism for non-experts to enter new instances. If concept definitions have to be changed, a new OWL file is generated from the installed ontology; it is transferred back to the editor, where it is changed, checked and finally re-installed.

The SR is designed to provide queries about static and slow-changing meta-data, e.g., information about available sensors, features of interest and their properties.

V. IMPLEMENTATION OF THE SEMANTIC REGISTRY

The Core implements the functionality of the SR. It uses an Ontology Store (OWLIM) for local storage of registered objects (concept instances and their specific relationships) and provides functions to manipulate and use the DSO.

The implementation provides several interfaces to access and manipulate the information as depicted in Figure 2.

A. Web based user interface (Web UI)

The Web UI comprises a form-based interface for data entry and an interactive search interface that enables users to browse through the TRIDEC meta-data using semantic search and to access other specific drill-down interfaces. It can be used to manually adjust imported data or as an alternative when an automatic data import is not possible.

Because an ontology can be quickly extended, e.g. with new sensor types, the SR must be flexible enough to support these new information types. Therefore, an important feature of the SR is that edit forms can be generated automatically from ontological concepts. By using such an edit form (see Figure 3); a user can improve the individual sensor descriptions registered automatically via e.g. the Message Oriented Middleware.
The semantic query editor (see Figure 4) allows a user to construct semantic queries without the knowledge of SPARQL. The generated SPARQL queries are editable by expert users. Combining the editor for individuals (see Figure 3) with this query editor (see Figure 4) allows an expert user to quickly prototype a complex sensor network and experiment with modifications before bringing them into production.

### B. SPARQL Endpoint

The most common query language for ontologies is SPARQL [11]. The SPARQL endpoint of the SR provides support for local and federated queries to other Semantic Registries and multiple knowledge bases. An example from the ISD scenario consists in selecting all data channels for a given well “Well01” of a given rig “InnovaRig”. The corresponding SPARQL query would be as follows:

```
PREFIX dso: <http://tridec.server/de/ontologies/TRIDEC5.2.2.owl#>
PREFIX ssn: <http://purl.oclc.org/NET/ssnx/ssn#>

SELECT DISTINCT (str(?mn) as ?mnemonic) (str(?u) as ?unit) WHERE {
  ?plat a dso:Rig .
  FILTER regex (str(?plat), "InnovaRig") .
  ?plat dso:hasProperty ?prop .
  ?plat dso:hasMnemonic ?mn .
}
```

This query will have the short names of the channels (mnemonic) as the result along with the corresponding units of measurement (unit).

### C. REST Interface

The RESTful [13] HTTP interface provides application developers with an easy-to-use programming interface that can be used to retrieve individuals of ontological concepts. The concepts are automatically mapped to the path of URI, e.g. the following REST URI would retrieve all sensors registered in the SR:

```
http://<SERVER_NAME>/rest/semreg/sensor/
```

The advantage of providing an easy access to the information without a required deep understanding of semantic technologies comes with the downside of losing the semantic power you have in the SPARQL interface.

### D. Message Oriented Middleware (MOM)

By providing access to the SR via a MOM, the component can be loosely coupled with systems. This makes it much easier to integrate into existing systems, which are part of a SoS. Currently, only the registering of new sensors is supported via the MOM by sending a message to the SR with the metadata of the sensor to add.

### VI. USAGE IN THE TRIDEC DRILLING SCENARIO

The implementation of the SR was strongly influenced by the requirements of the TRIDEC drilling scenario (ISD). The sensors mounted on several parts of the drilling platform must be able to publish their measurements for interested subscribers without interfering with each other. In terms of the MOM, which has been chosen for reliable communication, this means that the sensors have to use disjunct “topics” when publishing their data.

The process of registration of a sensor in the TRIDEC SR can be described as follows: the sensor uses one of the interfaces of the SR (REST or MOM) to transfer meta-data about itself to the registry. This information is structured as SensorML [14], following agreed profiles and guidelines. The SR parses this input and creates corresponding individuals to populate the Decision Support Ontology (DSO). By using this ontology as internal model of the SR, it is possible to use standard interfaces like SPARQL for querying the contents of the SR. As a side benefit, this model can be used for the generation of unique MOM topic names. The topic names are returned to the registrant of the sensor as the result of the registration request.

The consumers of measurement data may be interested in the SensorML description of the sensor publishing these data. The SR provides a REST command to retrieve the SensorML description based on the topic name. This second
use case reveals some of the difficulties in mapping SensorML to ontologies: The SensorML specification contains a multitude of alternative places for expressing the same facts. Therefore, a bijective mapping between different SensorML “styles” is almost impossible to achieve. The only guarantee after mapping back and forth is to get something semantically equivalent to the original SensorML. If – as in some cases – the original SensorML structure has to be preserved, it has to be stored inside the SR in addition to the ontology individuals generated from it.

VII. SUMMARY AND FUTURE WORK

In this paper, we presented the requirements for a geo-distributed SR in the context of a system-of-systems and important considerations for its design and implementation. A SR plays a central part in a SoS architecture that is able to fulfill the requirements A to D from section III. For example, in order to be able to implement Information Integration, the knowledge about available sensors and their data (e.g., geo-location, units, quality) is an indispensable prerequisite.

The benefits of the described solution are:

- The easy extendibility via the SSNO. For example, for managing descriptions of virtual sensors, such as model-based simulation services (for both scenarios), with the aid of their corresponding Sensor Processing Services (SPS) in the SR. This feature will especially support requirement E.
- The different available interfaces to the SR support different needs of expert users and systems, which could not be provided by a single access technology. Especially, the connectivity via a MOM supports the integration into existing systems by a message-based loose coupling of components.
- The web-based UI provides expert users with an experimentation platform for exploring new sensor types and networks.

On the downside, we identified the following issues:

- The SSNO is initially very high-level. A great deal of work is required for mapping domain-specific vocabularies and concepts to elements of SSNO and for adding the detailed sensor descriptions. This also involves many discussions with domain experts.
- In a distributed system, especially for critical systems, resilience and failover are hard requirements. Currently, the SR is a single point of failure. To overcome this limitation we are working on the resilient synchronisation of multiple geo-distributed SRs.

In the final phase of the TRIDEC project, we are focusing on missing or important features such as the management of selected types of observations and the querying of related meta-data and data via the MOM.

We are also considering the provision of a CSW interface (Catalogue Service for the Web [10]) to support harvesting of additional sources and for providing advanced search functionality to clients based on OGC standards.

ACKNOWLEDGMENT

Running from September 2010 to August 2013, TRIDEC is partially funded by the European Commission in its seventh Framework Programme under Grant Agreement No. 258723.

REFERENCES