A common data model for sensor network integration

Flora Amato, Valentina Casola, Andrea Gaglione, Antonino Mazzeo
Dipartimento di Informatica e Sistemistica
Universita' degli Studi di Napoli, Federico II
Naples, Italy
{*flora.amato,casolav,andrea.gaglione,mazzeo*}*@unina.it*

Abstract

One of the main open issues in the development of applications for sensor network management is the definition of interoperability mechanisms among the several monitoring systems and heterogeneous data. Interesting researches related to integration techniques have taken place, they are primary based on the adoption of sharing data-mechanisms; furthermore in the last years, the Service-Oriented Architecture (SOA) approach has become predominant in many sensor network projects as it enables the cooperation and interoperability of different sensor platforms at an higher level of abstraction. In this paper we propose a novel architecture for the interoperability of sensor networks, which is based on web services technologies and on the definition of a common data model enriched with semantic concepts and annotations. The proposed architecture allows the development of complex application by integration of heterogeneous data, accessible through services, according to standard data format and standard protocols.

1 Introduction

One of the main open issues in the multi-hazard approach to the environmental risk monitoring and management is the integration of heterogeneous data from different sources to manage and elaborate risk mitigation strategies, including emergency management planning. The large diffusion of sensor systems, together with their numerous applications has led to a huge heterogeneity in the logic for interfacing and collecting data from these systems.

However in order to automatize the elaboration of such heterogeneous data, specific integration frameworks for accessing different data sources are needed. Such frameworks should be able to access data, sensed by different sensing infrastructures; they should hide the heterogeneity among different sensor systems (in terms of sensor, networking, or middleware technologies) and provide a standard way to access them. Interesting researches related to integration techniques for heterogeneous sensor networks have taken place, but nowadays only few architectures have been proposed. Most of them try to define a common exchange mechanism among different sensor systems in order to facilitate the integration, and provide a software integration layer which allows different sensor systems to collaborate for the same purpose. Very often these solutions are tightly related to proprietary technologies and they are neither scalable nor open.

In the last years, the Service-Oriented Architecture (SOA) approach has become a cornerstone in many sensor networks projects. SOA-approach enables the cooperation and interoperability of different sensor platforms as it provides discovery, access and sharing of the services, data, computational and communication resources by the adoption of open standards. New open standards and the adoption of common data model to formally define and represent data knowledge, are the main features of these architectures that enable the definition of cooperative environments. As an example of data model, the OpenGeospatial Consortium provides an XML schema (the Sensor Model Language) for defining geometric, dynamic and observational characteristics of sensors and other standards as Observation and Measurement to describe observed phenomenon [1]. The standards: (i) provide general sensor information in order to support of data discovery, (ii) support the processing and analysis of sensor measurements, (iii) support the geolocation of the measured data, (iv) provide performance characteristics (e.g. accuracy, threshold, etc.), and (v) archive fundamental properties and assumptions regarding sensor. Within this context, the Campanian Region (Southern Italy) holds innovative environmental sensor networks to gather real-time and near-real time information on natural (i.e. seismic, volcanic, slope stability, pluviometric, oods and meteo-marine) and civil/infrastructural vulnerabilities. As a natural and densely populated lab for the multi-hazard problem, it is important for the Campanian region the development of an interoperability system to gather and elaborate heterogeneous data.

In this paper, a service based architectural model to provide interoperability among different sensor networks is proposed and a common data model for regional sensors is introduced and discussed in details. The strong advantage of web service architectures is to build information systems enabling the creation of applications by combining loosely coupled and interoperable services without the knowledge of the underlying systems. On the other hand, the common data model grants interoperability among multi-regional systems, providing a formal model for the integration of data gathered by different and heterogeneous sources. A formal data model specifies data relations, terminology and meanings. It is implemented through a set of ontologies and formally described in OWL (Web Ontology Language). As a result, we will present an innovative architecture for risk management that is completely based on services; it allows the integration of heterogeneous data and the implementation of applicative standard Web Services to let data (raw or aggregated) be available to authorized end-users and/or applications that need them. Then we will descrive a data collection service with the definition of a common data model for specific sensors. The analyzed data are related to seismic, volcanic, pluviometric and meteo-marine sensors. The model is applied in a first ”traslation step” and it is then converted according to the Observation and Measurement standard to grant interoperability. We will illustrate how the formalized data can be accessible trough a standard web services and can be invocated by other services for integration and elaboration purposes. The reminder of the paper is structured as follows: in Section 2 some related works are reported to assess the state of the art of many methodologies and technologies to face interoperability among heterogeneous data. In Section 3 we will present the architectural model by illustrating the main layers of the proposed architecture and their functionalities. In Section 4 we will illustrate the data model to enable interoperability. Finally in Section 5 some concluding remarks will be given.

2 Related Works

The need to guarantee interoperability among several monitoring systems and integration of heterogeneous data, can be seen from different point of viwes: *(i)* syntactic, to overcome technical heterogeneity; *(ii)*Semantic, to overcome ambiguities and different interpretations; *(iii)* Application, to deliver sustainable and re-usable concepts and components of the application domain; and *(iv)* Phenomena Observation, to evaluate the meaning of the observation from temporal, spatial and thematic perspectives. Some solutions to specific, but not complete, aspects of interoperability are available in the literature. The OGC [1] proposed a suite of specifications, named Sensor Web Enablement (SWE), to model sensor characteristics and services.

In particular, the suite includes: (i) Sensor Model Language(Sensor ML), (ii) Observation & Measurement and (iii) Sensor Observation Service. They allow to model sensor and sensor observations, data retrieval mechanism and web services (for access of the sensor data via web); it is possible to specify information as coordinates and timestamps, but they do not allow to state the semantics of the data and the meaning of sensor observations, making difficult the interoperability, the evaluation of the phenomena and the detection of situation awareness [3]. A first attempt of semantics definition for sensor Web has been proposed in the SSW framework [4], in which enhanced meaning for sensor observations is given by adding semantic annotations to existing standard sensor languages of the SWE, in order to increase interoperability and provide contextual information for situation awareness. Several efforts have been done in the data modeling field too: different ontologies for heterogeneous sensor data representation are proposed. An ontology presented by Ceruti in [10], models different concepts, as platforms and sensors, as characteristics tangible and intangible, and relationships and concepts such as data combinations. Another ontology for the sensor networks, presented by Eid et all in [11], provides semantic representation of sensor networks data, aiming at interpreting unambiguous structured information.

From the architectural interoperability point of view the current leading approach is based on Service Oriented Architectures (SOA). In [12] the authors introduce Web Service Resource Framework (WSRF) mechanisms into the core services implementation of the NICTA Open Sensor Web Architecture (NOSA). WSRF permit to perform simultaneous observational queries to heterogeneous Sensor Networks. Moreover, the GeoICT group at York University [13] have built an OGC SWE compliant Sensor Web infrastructure, developing a Sensor Web client capable of visualizing geospatial data, and a set of stateless Web Services called GeoSWIFT[14]. Furthermore, two main European early warning projects based on Service Oriented Architectures (SOA) have been proposed; the WIN (Wide Information Network) [6], [7] and ORCHESTRA [5] projects. WIN aims at developing an open and flexible platform to support multi-hazard and risk domains at European level, integrating national and regional data flows in the frame of a Web Service Architecture, and proposing a set of generic services and standard data modeling components that can be used in the deployment of several cases. The WIN Metadata Model is based on existing standards, such as Dublin Core, GML and ISO19115, and provides additional specifications for WIN. It permit to manage in uniform way all metadata by using the standard ebRIM(Electronic Business Registry Information Model). The WIN Data models are implemented via the Data Web Service component and the metadata via a ebXML Registry tool. Both components and tools use the standard SOAP/WSDL type of Web Services. An the other hand, ORCHESTRA adapts the ISO/IEC 10746 Reference Model for Open Distributed Processing to service-oriented architectures; adapting, in particular, the Reference Model for geospatial service networks on a process model compliant with the ISO standard RM-ODP, in order to design and implement geospatial SOAs. The ORCHESTRA architecture uses W3C Web services platform and the Geography Mark-up Language (GML) to implement web services, that are specified, as well as services, using UML diagram.

Different methods and techniques have been developed in order to assure the interoperability among several monitoring systems and integration of heterogeneous data. They represent solutions to specific interoperability and/or integration problems, but, at best of our knowledge, a comprehensive proposal of an architecture that provide a solution for the complete gamma of needs (i.e. able to manage phenomena observation, interoperability among several monitoring systems and integration of heterogeneous data, and that can be easily integrated with other pre-existing platform) does not exist, yet.

3 A Service-based Architecture for Sensor Network Integration

To design an open system and manage heterogeneous sensor data sources we propose a Service Oriented Architecture based on Web Services technologies. The adoption of standard technologies such as HTTP, XML, SOAP, WSDL, and UDDI [2] enables pervasive adoption and deployment of web services to reach interoperability. In fact, Web services are Internet based applications that communicate with other applications to offer business data or functional services in a standard fashion.

The proposed architecture offers several services and functionalities, which can be classified into two main categories: (a) services to elaborate data from heterogeneous sources; (b) services to manage all data sources (sensors, databases, simulators, etc.). We have located three horizontal layers: Data Collection Service Layer, Integration Service Layer, Application Service Layer, and some vertical layers whose services are transversal to the other layers, as Security, Management and Interoperability services, as illustrated in Figure 1. Each layer offers different services and is characterized by a deep functional specialization; furthermore, each layer is able to communicate with other layers through standard WSDL interfaces.

Grouping similar services in different layers offers different advantages: most complex services are built upon the collaboration of more elements of smaller complexities. Such approach concurs to build a more robust, scalable and maintainable architecture, with a better logical structure, too.

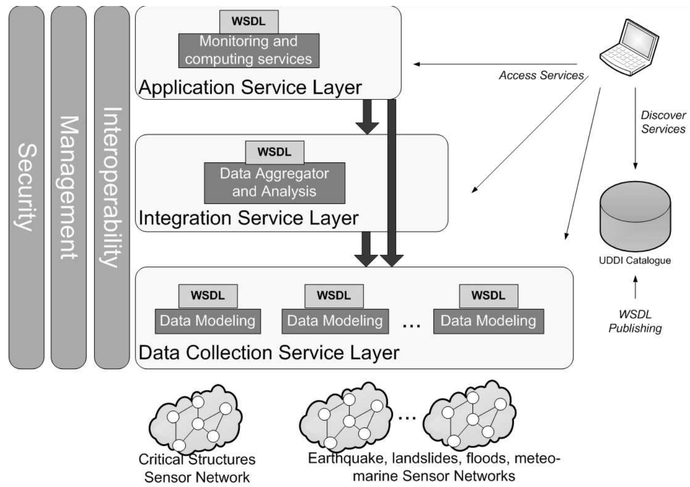

Figure 1. An oveview of the architecture

The structure in layers is a winning solution also from a security point of view as each service can have a specific security policy to control access and preserve data privacy. In the following, we provide a brief description of every layer and their specific services:

1. **Data Collection Service Layer:** in order to map the raw sensed data onto physical reality, and to provide the upper layer with a homogeneous view of the networks, a model of sensors, measurements and phenomena is required to complement the knowledge. The innovative approach that we propose aims to enrich sensor querying with ontological modeling techniques, mapping sensor data sets from a raw representation given by the sensor hardware/middleware to an XML/RDF data model and to an O&M standard representation. The Data Collection Service Layer provides the services to access heterogeneous sensor networks. Indeed, the services offered in this layer translate proprietary data format in a common data model at higher levels; translation is required in order to add ontological information and standard views as illustrated in next sections.

2. **Integration Service Layer:** while the Data Collection Service Layer is focused on interactions with a single sensor network, this layer contains services that are not associated with one specific resource but they are used to capture integration and interactions across collections of networks. For this reason, we refer to this layer of the architecture as the Integration layer. The services of this layer provide aggregated data sets

and numerical analysis values to the application level. It clusterizes network data sets according to the data model defined. Let us consider for example the case in which different technologies are adopted to sense the same data; while the collection layer is responsible to retrieve such data and represent them, this layer is capable of integrating data related to the same phenomena but observed by different heterogeneous networks.

3. **Application Service Layer:** different kinds of applications can be implemented for monitoring and elaborating complex data structures from heterogeneous sensor networks. They can implement warning threshold models or real-time event notification: moreover they can also simulate or foresee events. Applications are built by invoking and composing services defined in other layers; they all elaborate complex data structures formatted according to the data model defined and accessed via standard WSDL. Furthermore, each service has its own security policy and it is published in a public registry.

In Figure 1, every block represents available services; every arrow represents information flows (e.g. control and sensor data). All services are provided with a WSDL interface, they have been conceived to allow a standard interaction among modules. According to the WS standards, an interface between two components is a logical set of rules and data formats to allow an exchange of messages between components, based on a standard protocol (SOAP). Furthermore, every interface has been published in a UDDI catalogue. In this way every client/user can discover the services it needs and access them according to the access rights it has, through standard protocols. Thanks to this protocols and standards, the proposed architecture is able to grant a high level of interoperability and reveals a high degree of scalability while considering more different kinds of sensor networks.

We note that all data sets, which are stored and used to protect critical infrastructures, may be sensible data: this implies that all kind of private and sensible data have to be managed only by competent authorities.

The trasversal Security services enforce WS-Security standards to grant access (authentication and authorization) only to those authorized user. These services also deal with confidentiality and data integrity of the messages exchanged, non-repudiation of requests of messages and resilience to denial-of-service attacks.

In the following we will focus our attention on the Data Collection layer and, in particular, we will propose a common data model able to cope with the heterogeneity of different sensor networks.

4 Data Collection Service Layer

The Data Collection Service Layer aims at supplying the necessary mechanisms for the interoperability among heterogeneous sensor data and at providing meaning for sensor observations in order to enable situation awareness.

The basic problem to be addressed, when integrating data from heterogeneous sources, come from syntactic, schematic and semantics diversities of the schema. *Syntactic* heterogeneity refers to differences among paradigms used to format data such as plain text, relational db or xml documents. *Schematic* heterogeneity refers to different aggregation or generalization hierarchy defined for the same phenomena. Finally, *Semantic* heterogeneity regards disagreement on the meaning or interpretation on same data [8],[9].

Our proposal aims at facing and solving these differences. In particular, syntactic heterogeneity is resolved by converting proprietary data formats into XML document; schematic heterogeneity is resolved by adopting the standard O&M Language; finally semantic heterogeneity is resolved by defining a common data model formally described using an ontology.

The latter choice is motivated by the fact that many recent works on the use of ontologies showed an increase in the precision of service discovery queries when semantic representations were used over syntactic representations. In details, the Data Collection Service Layer provides the resources to access heterogeneous sensor network data and to translate proprietary data format into the defined common data model. To achieve this goal different modules are needed (see Figure 2):

1. Sensors add-on (data aggregation): this module is specific for each kind of sensor. It retrieves samples from sensor system and convert them into an XML format, according to the translation process defined for the specific network;

2. XML-RDF Wrapper: this module wraps the XML data sets, translating the file into an RDF document; an RDF validation and storage component is responsible for validating each RDF document against a domain ontology. If validation completes successfully data are saved in the storage system;

3. O&M Data modeler and Annotation: the last module retrieves an RDF document and converts it in an O&M standard XML document. The most important step is the annotation process: the O&M document is augmented with semantic information with an annotation process that extends the O&M standard with tags and attribute defined in our ontology.

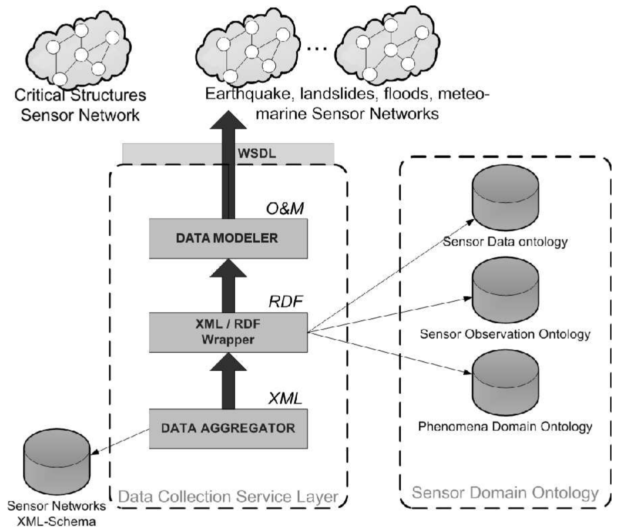

Figure 2. **The Data Collection Layer**

The **data aggregation module** acquires raw sensor data from heterogeneous sensor networks and encodes them into an XML document. The mapping is done using a common and extensible XML-schema defined for sensor networks and stored into the XML-Schema Repository.

The **XML-RDF Wrapper** translates the XML document into an RDF using the concept defined into the Sensor Data Ontology. The module basically works applying one or more XSLT according to the desired data format.

The **data modeler and annotation module** builds an O&M model from the data stored, adding concepts and relations between entities gained from reasoning on the sensor ontology. That information allows understanding the sensor information in order to eventually identify situation awareness. In this manner we add semantic annotations to an existing standard Sensor Web language in order to provide semantic descriptions and enhance the access to sensor data. This is accomplished with model-references to ontology concepts that provide more expressive concept descriptions. For example, by sing model-references to link O&M annotated *sensor data* with concepts within an OWL ontology allows one to provide semantics of sensor data; using a model reference to annotate *sensor device* ontology enables uniform/interoperable characterization/descriptions of sensor parameters regardless of different manufactures of the same type of sensor and their respective proprietary data representations/formats.

As example, we show a fragment of measurement data, codified in XML, and the corresponding "semantically annotated" fragment, codified in RDF and O&M :

Pluviometric Measurement Fragment in XML, RDF and O&M

```
<Sample>
 <Timestamp>2009-01-26T21:42:52</Timestamp>
 <Level>3.14159E0</Level>
</Sample>
```

```
<rdf:Description rdf:about="#PhysicalValues">
 <hasAccuracy
   xmlns="http://www.owl-ontologies.com/sensor.owl#"
   rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
   0.0
 </hasAccuracy>
 <hasDescription xmlns="http://www.owl-ontologies.com/sensor.owl#"
   rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
   Rain Level
 </hasDescription>
 <hasValue xmlns="http://www.owl-ontologies.com/sensor.owl#"
   rdf:datatype="http://www.w3.org/2001/XMLSchema#float">3.14159</hasValue>
 <hasUnit xmlns="http://www.owl-ontologies.com/sensor.owl#">
  <rdf:Description xmlns="http://www.owl-ontologies.com/sensor.owl#"
    rdf:about="http://sweet.jpl.nasa.gov/2.0/sciUnits.owl#millimeter"/>
 </hasUnit>
</rdf:Description>
```

```
<om:result xsi:type="gml:MeasureType" uom="urn:ogc:def:uom:OGC::mm">
   3.14159
</om:result>
```

In the next section, we will give some details on the proposed RDF data model, used to guarantee interoperability among heterogeneous sensor networks.

4.1 The Common Data Model

The proposed RDF common data model is made of two main classes: the **Sensor Class** includes knowledge about sensor specification, as identification, type, characteristic, temporal and spatial attributes; the **Measurement Class** defines information about the data collected by the different sensors. As illustrated in Figure 3(a), the Sensor Class models the sensor characteristic such as $Location$, $Network$, $Measurement$, $SensorType$. Among the others, the $Location$ entity describes geographic information by using coordinates in DMS or UTM system, and is compliant to the WGS84 standard[1]. The $SensorType$ entity specifies information about typology description.

As illustrated in Figure 3(b), the Measurement Class contains information about physical characteristics of gathered measured samples, every measurement is characterized by a set of physical parameters, as the $AcquisitionTime$ that indicates the temporal interval of phenomena observations, and the $SamplePosition$ that indicates the position of the samples at the beginning of the observation, specified in the same format of the position of the sensors itself. Such physical parameters are supplied by statistic information (as covariance and main values), sensor parameters (as sensor power, voltage, bearing and direction), and provide, beyond the value itself, the accuracy and the measurement unit (that can be defined in external ontologies[2], too).

[1]The World Geodetic System is a standard for use in cartography, geodesy, and navigation. It comprises a standard coordinate frame for the Earth

[2]http://sweet.jpl.nasa.gov/2.0/math.owl

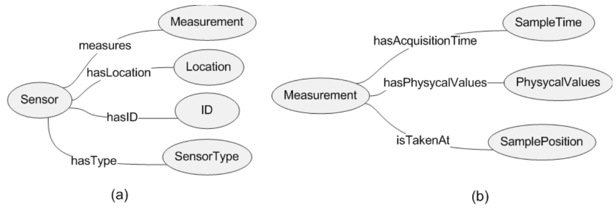

Figure 3. The Sensor and Measurement Classes

Furthermore, we can refer to external ontologies to add further concepts and relations, to model not only the sensor network but also the Observation and the Phenomena. For example, the sensors can be grouped in classes, associated to a specific phenomena domain ontology (let us consider for instance, a temperature sensor, a humidity sensor and a wind sensor that may collectively monitor the same phenomena, they may refer to the same weather sensor class). Finally, thanks to this model, it is possible to share and standardize, at a higher level in the service, the data format and so resolve possible conflicts generated by heterogeneous sources. The defined ontologies are needed to give a shared semantic that describe the sensor data and observations in unambiguous way.

Note that such semantic information can not be derived by analyzing the sensor data or specified using the SWE languages; those standards, in fact, allow to characterize physic concepts, such as spatial and temporal coordinates, but not to specify the meaning of sensor observations, that are necessary, in order to allow the interoperability of heterogeneous sensors and improve situation awareness.

5 Conclusions and Future Works

In this paper we have proposed an innovative architecture for the interoperability of sensor networks; it is based on web services technologies and on the definition of a common model for enabling semantic-based data management in sensor networks. The model promotes interoperability in presence of heterogeneous sensors, data and applications. The formal modeling of data also allows for events analysis and prediction, exploiting reasoning techniques on semantic relevant information. The division between data format and semantics, makes the proposed model general enough to face changes in sensors technologies, communication standards, or sensors networks topologies.

References

[1] D. Gay, P. Levis, D. Culler, E. Brewer, “nesC 1.1 Language Reference Manual, http://nescc.sourceforge.net/papers/nesc-ref.pdf”, March 2003.

[2] D. Booth et al. , Web Services Architecture, W3C Working Group, February 2004 http://www.w3.org/TR/ws-arch

[3] A. Sheth and M. Perry, Traveling the Semantic Web through Space, Time, and Theme, IEEE Internet Computing, vol. 12, no. 2, 2008, pp. 81 86.

[4] A. Sheth and C. Henson, S. Sahoo, Semantic Sensor Web, IEEE Internet Computing, vol. 12, no. 4, 2008, pp. 78 83.

[5] X. Chu and R. Buyya, Service Oriented Sensor Web, Sensor Networks and Configuration (2007), pp. 51 74.

[6] C. Alegre, C. Monfort, P. Lazaridis. WIN : a new Service Oriented Architecture for risk management. www.vfdb.de/riskcon/paper/Interschutz Conference-WIN Paper.pdf

[7] C. Alegre, H. Sassier, S. Pierotti, P. Lazaridis. A New Geo-Information Architecture for Risk Management, Springer - Geo-information for Disaster Management, 2005, pp. 543-550.

[8] F. Hakimpour, A. Geppert, Resolving semantic heterogeneity in schema integration, Proceedings of the international conference on Formal Ontology in Information Systems, p.297-308, October 17-19, 2001, Ogunquit, Maine, USA.

[9] Natalya F. Noy, Semantic integration: a survey of ontology-based approaches, ACM SIGMOD Record, v.33 n.4, December 2004 .

[10] M. G. Ceruti, Ontology for Level-One Sensor Fusion and Knowledge Discovery. Knowledge Discovery and Ontologies Workshop at ECML/PKDD, Pisa, Italy, September 2004.

[11] M. Eid, R. Liscano, A. El Saddik. A Novel Ontology for Sensor Networks Data. CIMSA 2006 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications La Coruna - Spain, 12-14 July 2006

[12] T. Kobialka, R. Buyya, C. Leckie, R. Kotagiri. A Sensor Web Middleware with Stateful Services for Heterogeneous Sensor Networks Intelligent Sensors, Sensor Networks and Information, 2007, ISSNIP 2007.

[13] Sensor Web: http://sensorweb.geomatics.ucalgary.ca/

[14] Liang S, Coritoru, A., Tao, C. A Distributed Geo-Spatial Infrastructure for Smart Sensor Webs, Journal of Computers and Geosciences Vol.31, 2005.