Knowledge-based service architecture for multi-risk environmental decision support applications

Stuart E. Middleton and Zoheir A. Sabeur

IT Innovation Centre, University of Southampton, Gamma House,
Enterprise Road, Southampton SO16 7NS, United Kingdom
{sem, zas}@it-innovation.soton.ac.uk

Abstract. This paper describes our work to date on knowledge-based service architecture implementations for multi-risk environmental decision-support. The work described spans two research projects, SANY and TRIDEC, and covers application domains where very large, high report frequency real-time information sources must be processed in challenging timescales to support multi-risk decision support in evolving crises. We describe how OGC and W3C standards can be used to support semantic interoperability, and how context-ware information filtering can reduce the amount of processed data to manageable levels. We separate our data mining and data fusion processing into distinct pipelines, each supporting JDL inspired semantic levels of data processing. We conclude by outlining the challenges ahead and our vision for how knowledge-based service architectures can address these challenges.

Keywords: OGC, SWE, SOA, semantics, data fusion, environmental risk management, crisis management

1 Introduction

Multiple environmental risks, including those leading to crisis events, require fast and intelligent access to relevant spatial-temporal environmental information with meaningful thematic context by decision makers. The potentially very large and heterogeneous information generated from such data sources should be critically integrated and coherently presented alongside uncertainty information such as sensor accuracy or modelling error estimations. The handling of such information complexity requires agile and information channelling, supported by intelligent data filtering, mining and fusion.

This paper describes our work to date on environmental service oriented architecture implementations for multi-risk environmental management and decision-support. The work described spans two projects, SANY [8] and TRIDEC [10], and covers application domains where very large, high report frequency information sources are processed in challenging timescales in order to support multi-risk decision making in evolving crises.

Our approach to implementing knowledge-based service architectures is to use standards from the Open Geospatial Consortium (OGC) and World Wide Web Consortium (W3C). These support semantic interoperability, allowing metadata driven automation when integrating new data sources. We also propose the use of the Business Processing Execution Language (BPEL) to achieve agility for our data

processing services, and the identification of clear semantic levels for each data fusion & mining processes under a structured Joint Directors of Laboratories (JDL) type data fusion framework.

Finally, we propose context-ware information filtering methods to intelligently filter and index raw information events and tailor processing.

2 Problem statement

We now live in an information age with increasing volumes of information from affordable means of communication, monitoring and observation systems. This information is more accessible to much larger communities of multi-disciplinary users than ever before. Such large volumes of data, and associated high reporting event frequencies, require data to be stored, intelligently retrieved, analysed and efficiently distributed to groups of collaborating users.

The domain of air quality monitoring, marine risk management and geo-hazard risk management in urban areas has been examined in the SANY project [8]. These applications areas are characterised by their use of in-situ sensors & sensor networks, remote sensing data and contextual information. Decision timescales range from minutes to days, with tens of sensor stations reporting measurements and dataset sizes in the gigabyte range. For example, in the area of bathing water quality risk management, beach attendants need to receive alerts about microbial contamination risks levels of exceedance under the EC bathing water directives. Exceedence levels are predicted from a fusion of meteorological and hydrological sensor measurements fed into simulation data models.

The domains of natural crisis management and industrial drilling operations are being examined within the on-going TRIDEC project [10]. These domains are characterised by having survey data sizes in the terabyte range, which are likely to increase to petabytes of data in the next decade. These application domain areas use a variety of information and data sources, such as sensor networks with hundreds of sensors, camera streams, textual data and social networking web 2.0 sites.

While oil drilling operation decision-making timescales range from hours, for planning decisions, to few seconds for detecting drilling operation system anomalies, the intergovernmental agencies decision timescales, for tsunami warning, range from an hour down to few minutes, depending on the distance a tsunami will travel before impacting on the coastline.

3 Developing a knowledge-based service architecture

Semantic interoperability within communities, and also between traditionally disparate communities is a major challenge to overcome. The W3C community has developed standards for the semantic web such as XML, RDF and OWL. The OGC [5] SWE standard set [9] has also been developed to handle different sensor types (in-situ, remote sensors, video, models, etc.) from a variety of different disciplines such as those in environmental sciences, defence, crisis management, and spatial

application domains covering marine, atmospheric, terrestrial biodiversity and so forth.

SWE is a suite of OGC standards consisting of three standard XML encodings (SensorML, O&M, SWE Common) and four standard web service interfaces (SOS, SAS, SPS, WNS). The SWE standards are predicated on a service oriented architecture (SOA) approach. UncertML is an emerging standard [11] for representing statistics and different mathematical distribution types that the SWE working group are currently discussing with a view to integration into their next generation of standards. It can be seen in Figure 1 that standards now exist to support knowledge-based services working in a variety of different domain layers.
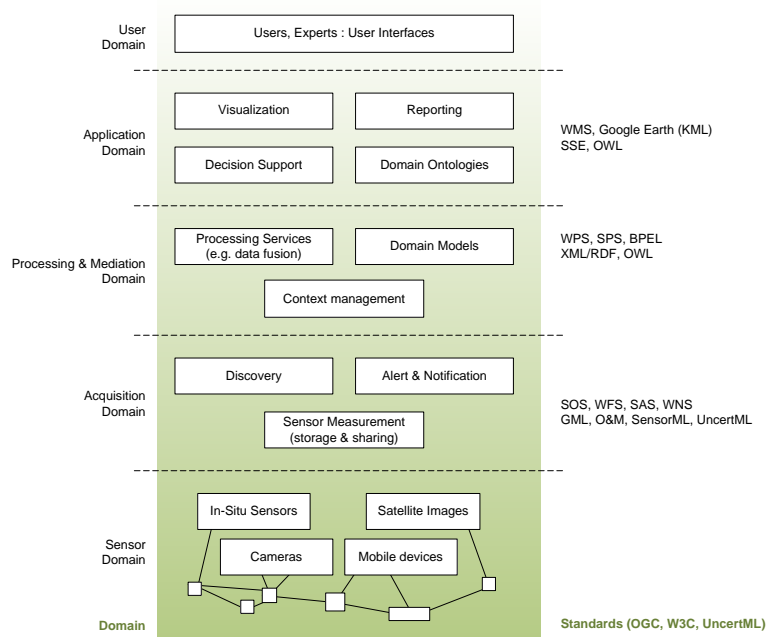

Fig. 1. Relationships between different domain layers and available standards.

As sensor systems become ever larger and more inter-connected we are seeing the volumes of data being shared moving from gigabytes to petabytes (especially in satellite systems [2]). In these circumstances, the need for context-aware data filtering and processing becomes important for achieving critical information management and decision-support. By using different types of context (e.g. a decision support task context or trust model in a data source) intelligent information filtering algorithms can reduce the volume of relevant data which needs to be processed. Context-aware algorithms can also improve the effectiveness of data mining and data fusion approaches by helping to steer algorithms according to the temporally evolving needs of decision makers in a crisis.

The underlying work on knowledge-based service architecture, through projects such as SANY and TRIDEC, consists of identifying semantically well-defined levels of data fusion, motivated by JDL data fusion information model [3]. The separation of

the semantic layers involved in data processing allows us to implement structured and specialized processing units that can be orchestrated into agile processing pipelines, which could be well suited to respond to the constantly changing requirements of multi-risk environmental decision support applications. These semantic layers can be seen in Figure 2.
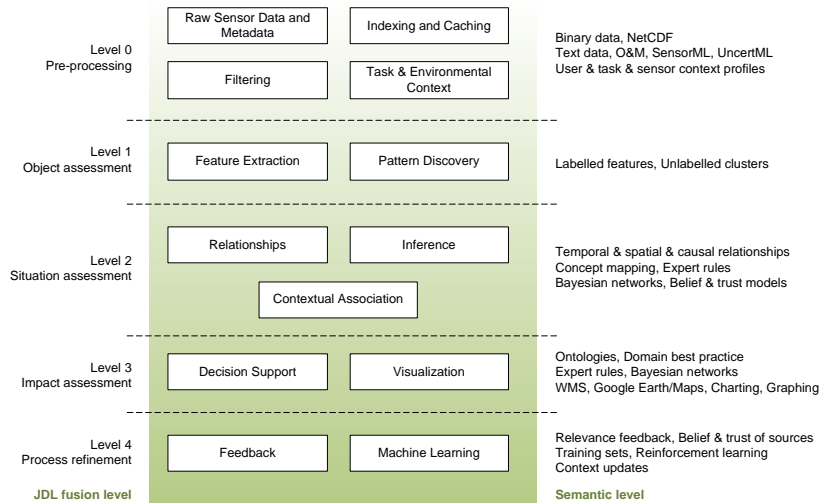

Fig. 2. Mapping semantic layers to structured levels of data fusion

Existing environmental information systems, such as those implemented in the OSIRIS [6] and ORCHESTRA [12] projects, make use of OGC standards and service orchestrated architectures. Other relevant projects like GITEWS [7], and its follow-on DEWS [1], provide examples of open sensor platforms, using OGC standards to integrate sensor networks. These projects use OGC semantic metadata to assist semantic interoperability, however, they do not address the issues of handling large volumes of data. Also, they provide limited agility in response to the changing processing requirements of dynamic decision support situations that occur in risk management. The novel approach which we aim to adopt is the coupling of de-facto data fusion methodologies, state of the art scalable processing architectures and semantic technology in order to overcome the above mentioned shortcomings.

4 Knowledge-based services for in-situ sensors & sensor networks

The SANY project [8] focused on interoperability of in-situ sensors and sensor networks, assuring sensor data could be easily processed and used as a basis for decision support. The sensor service architecture (SensorSA) is the fundamental architectural framework of the SANY project for the design of sensor-based environmental applications and their supporting service infrastructure.

The SensorSA is a service-oriented architecture (SOA) with support for event processing and a particular focus on the access, management and processing of

information provided by sensors and sensor networks. The foundation for the SensorSA is the ORCHESTRA project architecture (RM-OA) [12] and the OGC SWE architecture. The knowledge-based service architecture in SANY is a concrete example implementation [4] of the SensorSA architecture principles. This is shown in Figure 3.
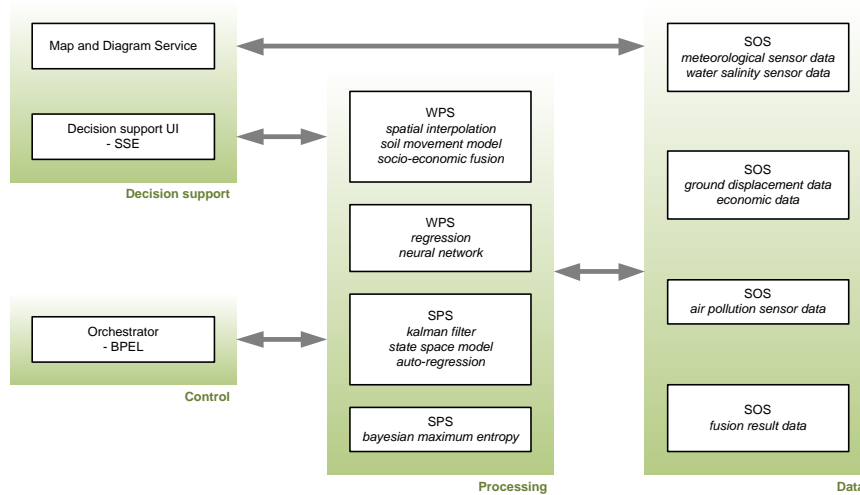

Fig. 3. Knowledge-based service architecture [SANY]

All sensor data in SANY was obtained directly from a sensor observation service (SOS). In SANY SOS's provided access to both sensor measurement datasets and fusion processing result sets. Sensor datasets contained measurements for air pollution, meteorology, ground displacement and more. Whenever a web processing service (WPS) or sensor planning service (SPS) needed data it used a SOS client.

Because sensor data was represented using both the SWE observation and measurement (O&M) model and UncertML, sensor metadata (units, measurement types, sensor accuracy etc.) was directly available to be used by the SANY processing services. Much of SANY's data fusion pre-processing was thus automated, using metadata to aggregate data from separate distributed SOS's. Examples of this are the use of metadata to identify identical phenomenon for dataset merging and the use of unit metadata to validate data value ranges. In SANY new sensor datasets from SOS's could just be 'plugged in'.

An example of the type of data processing work performed in SANY is spatial interpolation of meteorological data, for subsequent input into a bathing water quality data-driven model. The semantic level of fusion in SANY was mostly a combination of pre-processing (level 0) and impact assessment (level 3).

For complex multi-service workflows in SANY a BPEL orchestrator service was used to execute workflows involving SOS's, WPS's and SPS's. The results of these workflows would be sent to the decision support services, making use of geospatial mapping services and advanced visualizations (e.g. Google Earth 3D visualizations). In this way, raw sensor data, data fusion results and the uncertainty context associated with this data could be made available to decision makers.

5 Knowledge-based services for real-time & multi-modal data

In the on-going TRIDEC project [10] we are building on the approaches developed within in SANY, focussing on the challenges associated with handling very large scale multi-modal data sources in real-time. In common with SANY we are using OGC and W3C standards to achieve semantic interoperability. To achieve scalability and performance we will employ a message oriented middleware (MOM), allowing sensor measurements to be sent as events on a message bus. An example of an event is a set of measurements taken over a sample period. Using a message bus allows us to employ complex event routing, and dynamically configure this routing based on the current decision support requirements and context. Context-aware filtering of events, as described in figure 4, is essential as we are planning to handle up to a thousand multi-modal data sources, each source measuring in real-time tens of properties with sampling at periods down to the millisecond range. It is simply not possible to naively receive and process this amount of unfiltered data within the decision making time window of our applications.
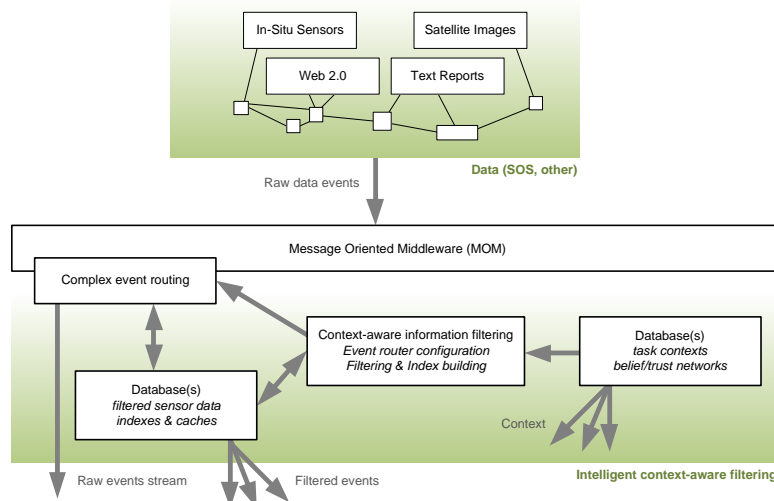

Fig. 4. Illustration of intelligent context-aware information filtering [TRIDEC]

Because crisis situations are dynamically changing, the decision making task context is modelled in real-time. Up-to-date task context allows the knowledge-based services' processing framework to re-configure itself in an agile fashion, through re-focussing of processing services to answer the questions required by decision makers at any given evolutionary stage of a crisis. OGC SWE services, such as sensor planning services (SPS's) and web processing services (WPS's), can be used to control the specific processing steps in a number of processing pipelines. A BPEL orchestrator can enact specific pipeline workflows. However, it will need to be controlled by a choreography component dynamically linking the questions decision makers need to answer to known pipelines capable of generating the answers.
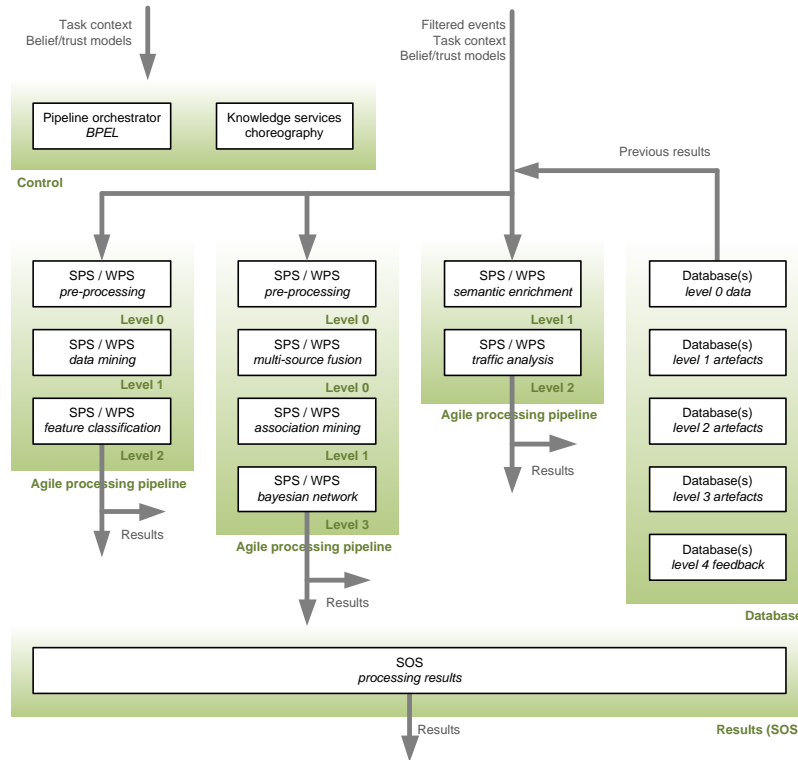
Fig. 5. Illustration of an agile processing framework for knowledge-based services [TRIDEC]

Each processing step provides result data at a different semantic level. An example of a processing step would be feature extraction, annotating a fused multi-sensor data feed with metadata to describe temporally correlated patterns. Each processing pipeline provides the processing steps needed to build up evidence that is able to answer a specific 'question' of interest to a decision maker. An example of a pipeline would be a set of several processing steps generating reports, in real-time, on sensor correlations matching conditions where a drilling 'kick' looks likely to occur.

The use of multi-modal data sources is a challenge in itself. Sensor data, text, video and image data all present their own processing requirements. Our agile processing framework allows us to include semantic enrichment processing steps into our pipelines, facilitating annotation of real-time information feeds with metadata suitable for multi-level semantic processing.

6 Conclusions

Environmental information systems are becoming more and more complex as they increase in scale and scope. As sensor systems become more advanced, the data volumes also increase, both in terms of data sizes (petabyte datasets) and message

throughput (hundreds of sensors reporting measurements multiple times per second). These data volumes are too large to be processed naively, and are forcing us to develop context-aware information filtering capabilities into our knowledge-based service architecture.

We are starting to see federated approaches involving the integration of legacy sensor networks and processing capabilities, coupled together in system of systems type architectures. Such federated systems increasingly require the use of standards, such as the OGC and W3C standards sets, to facilitate semantic interoperability and allow seamless automated data processing across aggregated distributed sensor networks.

We are also seeing time-critical decision support environments, such as natural crisis management and industrial drilling operations, which present challenging decision time scales ranging from hours to seconds. This type of multi-risk decision support requires an agile real-time processing architecture which is capable of responding to changing decision support requirements as the crisis unfolds over time.

Our work in projects like SANY and TRIDEC is allowing us to experiment with architectures that couple context-aware information filtering, agile processing and context-aware data fusion and mining in a structure framework.

References

1. DEWS project, Distant Early Warning System, FP6-045453, http://www.dews-online.org
2. Gehrz, R.D.: The NASA Spitzer Space Telescope, Review of scientific instruments 78, 011302 (2007)
3. Lambert, D. A.: A blueprint for higher-level fusion systems. Information Fusion, 10 (1), 6-24 (2009)
4. Middleton, S.E. (eds.): SANY fusion and modelling architecture, OGC discussion paper, OGC ref 10-001 (2010)
5. Open Geospatial Consortium (OGC), http://www.opengeospatial.org/ogc
6. OSIRIS project, Open architecture for Smart and Interoperable networks in Risk management based on In-situ Sensors, FP6-0033475, http://www.osiris-fp6.eu/
7. Rudloff, A., Flueh, E. R., Hanka, W., Lauterjung, J., Schöne, T.: The GITEWS Project (The German-Indonesian Tsunami Early Warning System), 3rd General Assembly European Geosciences Union (2006)
8. SANY project, Sensors Anywhere, FP6-IST 0033564, http://www.sany-ip.eu
9. Sensor Web Enablement (SWE), http://www.opengeospatial.org/projects/groups/sensorweb
10. TRIDEC project, Collaborative Complex and Critical Decision Support in Evolving Crises, FP7-258723, http://www.tridec-online.eu
11. Uncertainty Markup Language: UncertML, v1.0.0, http://www.uncertml.org
12. Uslander, T. (eds.): RM-OA: reference model for the orchestra architecture, OGC best-practice document, OGC ref 07-097 (2007)