On-the-fly Sensor Fusion for Real-time Data Integration

Bernd RESCH [a,b,c,1]
[a] *University of Osnabrück*
[b] *Research Studios Austria*
[c] *Massachusetts Institute of Technology*

Abstract. A key requirement for performing meaningful geographic information analysis processes is high quality of the underlying data. Quality criteria for geo-data comprise accuracy, completeness and up-to-dateness. While the first two parameters have recently received a lot of attention, the up-to-dateness parameter is still widely neglected, but is rapidly gaining importance. Thus, a mechanism for integration of sensor measurements and other live data has to be created, which combines different data sources in real time. This paper presents an on-the-fly sensor fusion method including its implementation. The proposed solution shows several advantages versus existing approaches such as elimination of bottlenecks, optimised workflow efficiency, as well as broad interoperability on data and on service levels.

Keywords. sensor fusion, on-the-fly data harmonisation, open-source, sensor web, sensor observation service, data homogenisation, data fusion.

1. Introduction

A key requirement for performing meaningful Geographic Information (GI) analysis processes is high quality of the underlying data [1]. Quality criteria for geo-data comprise accuracy, completeness and up-to-dateness. While the first two parameters have recently received a lot of attention, the up-to-dateness parameter is still neglected in many cases. This is partly due to the fact that GIS data is per definition historic and so far, geospatial analysis has focused on processing static data, with low temporal fluctuations such as demographic data, remote sensed images, socio-political data, digital elevation models (DEM) or different kinds of infrastructure-related data. With the emergence of a variety of real-time data sources (e.g. sensor networks, geo-referenced cameras, Radio Frequency Identification [RFID] based systems etc.), and with the creation of the new paradigm of 'real-time geography' and GI motivated situational awareness, topicality of base data is now rapidly gaining significance.

These developments necessitate the creation of a mechanism for real-time integration of sensor measurements and other live data to combines different data sources in real time and to provide those data via standardised interfaces such as OGC Web Feature Service (WFS) or Web Map Service (WMS). This allows for querying heterogeneous sensor data in real time and their integration into GI analysis systems.

[1] Corresponding Author: Visiting Professor at University of Osnabrueck, Barbarastrasse 22b, 49076 Osnabrueck. Email: bernd.resch@uos.de

2. Disambiguation of the term 'Sensor Fusion'

As their definitions vary, the terms *sensor fusion*, *data fusion* and *geo-processing* overlap to a certain degree in the sense that they all combine different kinds of data sources to achieve a meaningful outcome by applying various types of algorithms.

It is striking that data fusion definitions mostly do not use the term *processing* at all, but that geo-processing definitions distinguish clearly between the two terms (e.g. the ISO 19119 standard). This also results in the fact that there is no consistent and uniformly applicable definition for both terms. Considering the goals and application scenarios of the GI research field, the usage of the terms as follows can be defined as follows:

- *Sensor Fusion*: the enrichment of raw measurements towards 'entities' so they can be used as input 'datasets' for further processing steps
- *Data Fusion*: combination of several data sources with congeneric content (e.g. same data structures, objects, features etc.). This also includes basic data manipulation functionality (e.g. averaging, calculating mean values, simple filtering etc.).
- *Geo-Processing*: applying geospatial analysis algorithms on geo-referenced data to create meaningful and use case tailored information layers.

Previously, the term sensor fusion has be defined as 'determining that the data from two or more sensors correspond to the same phenomenon' [2], amongst others. Thus, it shall be stated that besides real-time sensor measurements, the data fusion and the geo-processing steps will most likely also integrate legacy geospatial data like road networks, census data, or topographic models, and may also include historic/archived measurements.

The output of the different processes can be summarised as follows. The result of a *sensor fusion* process is raw data, e.g. the Observations and Measurements (O&M) [3] output of SOS data – harmonised data from several sensors of the same kind. The output of *data fusion* is data, e.g. the results of combining of several data sources, i.e. sensor fusion results, such as CO_2 and particulate matter concentrations combined from in-situ and Earth Observation (EO) sensors for different geographic areas. This also includes calibration and simple statistical analysis, i.e. mean value over time, variances, maximum and minimum values. On the contrary, *geo-processing* output can be seen as enriched information layers, i.e. analysed data.

3. Related Work

Ziegler and Dittrich [4] mention that the central question in real-time data integration research is how to combine data to offer the user only one view on these data. This requires progress as well in semantics (reduction of losses, common object descriptions etc.) as in data modelling concepts. The authors describe the issues solely from a database perspective.

The same applies to Rittman who presents Oracle's approach [5], which basically consists of a middleware between (web) services and a continuously updated database. Like Sybase Inc. [6], the Oracle approach is able to detect database events in order to analyse heterogeneous data sources and to trigger actions accordingly. Rahm et al. [7]

developed a more dynamic data fusion approach by suggesting the use of on-the-fly object matching and metadata repositories to create a flexible data integration environment.

These approaches for real-time data integration rely on the costly step of creating a temporal database, which is only suitable in a powerful computation environment without strict performance constraints.

Harrie [8], and Lehto and Sarjakoski [9] present data integration web services based on the classic request/response model. Although both methods widely use open geospatial standards, they are not suitable for the integration of real-time data for large volumes of data.

Sarjakoski et al. [10] introduce a real-time spatial data infrastructure (SDI), which performs several application-specific steps (coordinate transformation, spatial data generalisation, query processing or map rendering and adaptation), but accounts neither for event-based push mechanisms nor for the integration of sensor data.

A number of approaches try to integrate sensor measurements in real time. However, Srivastava et al. [11] and Balazinska et al. [12] present very general concepts in a data integration architecture approach describing the data fusion process chain, but there are no concrete conclusions how the final goal of establishing such an infrastructure could be achieved. A more technical approach addressing measurement data originating from ad-hoc sensor networks is described by Riva and Borcea [13]. The authors discuss challenges to make heterogeneous sensor measurements combinable by creating highly flexible middleware components. The approach is greatly application-motivated and thus very detailed as far as specific implementation details are concerned, but does not offer a generalised model for real-time data integration.

Finally, there are several application-driven data fusion approaches, which try to integrate different kinds of (spatial) data in a very specific way. Different application areas are addressed including financial data integration [14], integrating Corporate Performance Management (CPM) and Business Intelligence (BI), clinical data integration ([15], [16]) or heterogeneous data visualisation ([17]). However, these approaches do not focus on the use of open standards and they are not considering the combination of push mechanisms and real-time data integration.

4. Sensor Fusion Service Implementation

Originating from the shortcomings of the previous approaches described in section 3, the Sensor Fusion Service (SFS) presented in this paper has been conceived and implemented. It constitutes the connecting component between the geo-sensor web (i.e. Sensor Observation Service based sensor networks) and the more widely adopted WFS/WMS interfaces. The direct conversion of data structures allows the easy integration of sensor data into Geographic Information Systems (GIS) based applications and therefore enables fast and ubiquitous data visualisation and analysis.

From a technology viewpoint, the Sensor Fusion Service acts as a ‘translator’ between O&M-encoded measurement data and the provided OGC WFS/WMS output. During the transformation procedure, certain input parameters (spatial reference system, units of measure, data structures etc.) are interpreted. Figure 1 illustrates the general functionality of the Sensor Fusion Service, which is realised as a custom data store for the open-source *GeoServer* [18], currently in version 2.1.

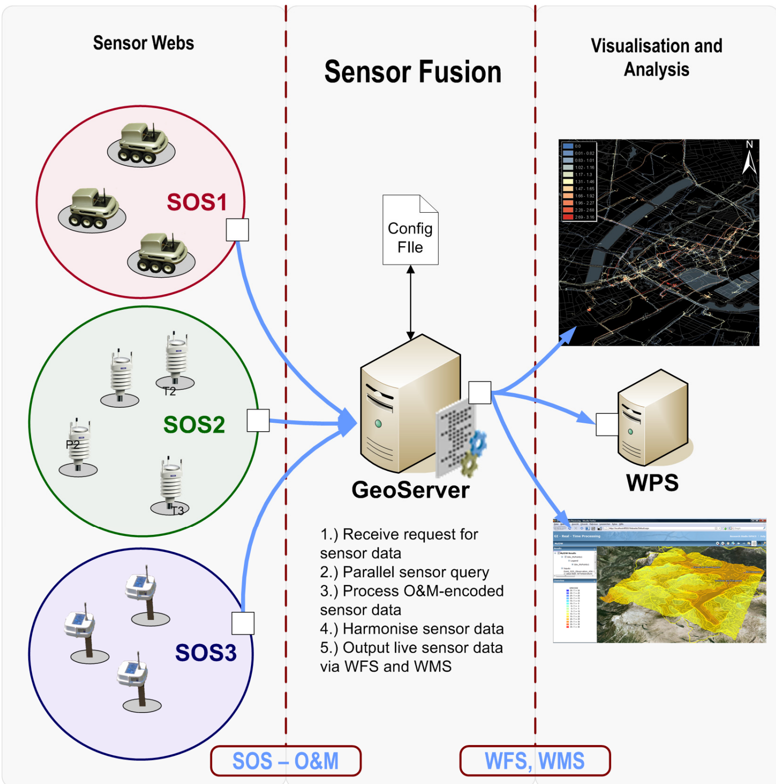

Figure 1: Sensor Fusion Service (SFS) General Functionality.

This plug-in datastore enables the direct integration of OGC SOS responses (in O&M [Observations and Measurements] format) into GeoServer and their conversion to other OGC-conformal service messages on the fly.

Table 1 illustrates the data flow sequence within the data store for a feature/observation query in the form of method calls between the different classes of the data store. The sequence is divided in five segments. GeoServer's WFS/WMS services are the external output interface towards the end user side, whereas the *SOS2ObservationList* class is the external interface towards the sensor networks.

Table 1. Segments in the Sequence Diagram for the Sensor Fusion Query.

Segment	***Description***
1	GeoServer requests the 'schema' of the queried feature type. Like this, GeoServer gains knowledge about the data structure – meaning the contained elements and their data types – and can thus dynamically build an accordant data query.
2	GeoServer creates a *Datastore* object, which is later used for the actual data query.
3	The data store calls the *getObservationList()* method, which queries the sensors in a parallel manner by making use of multi-threading technology. The measurement values are then assembled in a combined *ArrayList*, which is returned to the data store.
4	For the actual operation of processing sensor data, the data store generates a *FeatureReader* object, which again creates an *AttributeReader* object. The latter reads the measurement data sets one by one, disassembles the single attributes according to the data schema, which has been requested in step 1, and saves them in the *FeatureReader* object.
5	The data store creates a *FeatureCollection* from the attribute data obtained in step 4 and returns it to GeoServer. From this data structure, the server is able to produce a variety of different output formats, such as WMS images, geographically enables Really Simple Syndication (GeoRSS), Keyhole Markup Language (KML), Geography Markup Language (GML) and numerous more.

As described above, the O&M data store for GeoServer constitutes the connecting component between the geo-sensor web (Sensor Observation Service) and the standardised WFS/WMS interface, which is visible toward the internet for external clients. Figure 2 illustrates the data flow sequence within the data store for a feature query in the form of method calls between the different classes of the data store. The sequence diagram is divided in five segments, as described in Table 1.

The classes *SOSClientConfigFileHandler* and *SOS2ObservationList* are responsible for the actual data query, i.e. the communication with the external geo-sensor network, whereas the other classes serve for integrating sensor data into GeoServer and for transforming O&M-encoded measurement data into WFS-conformant output formats.

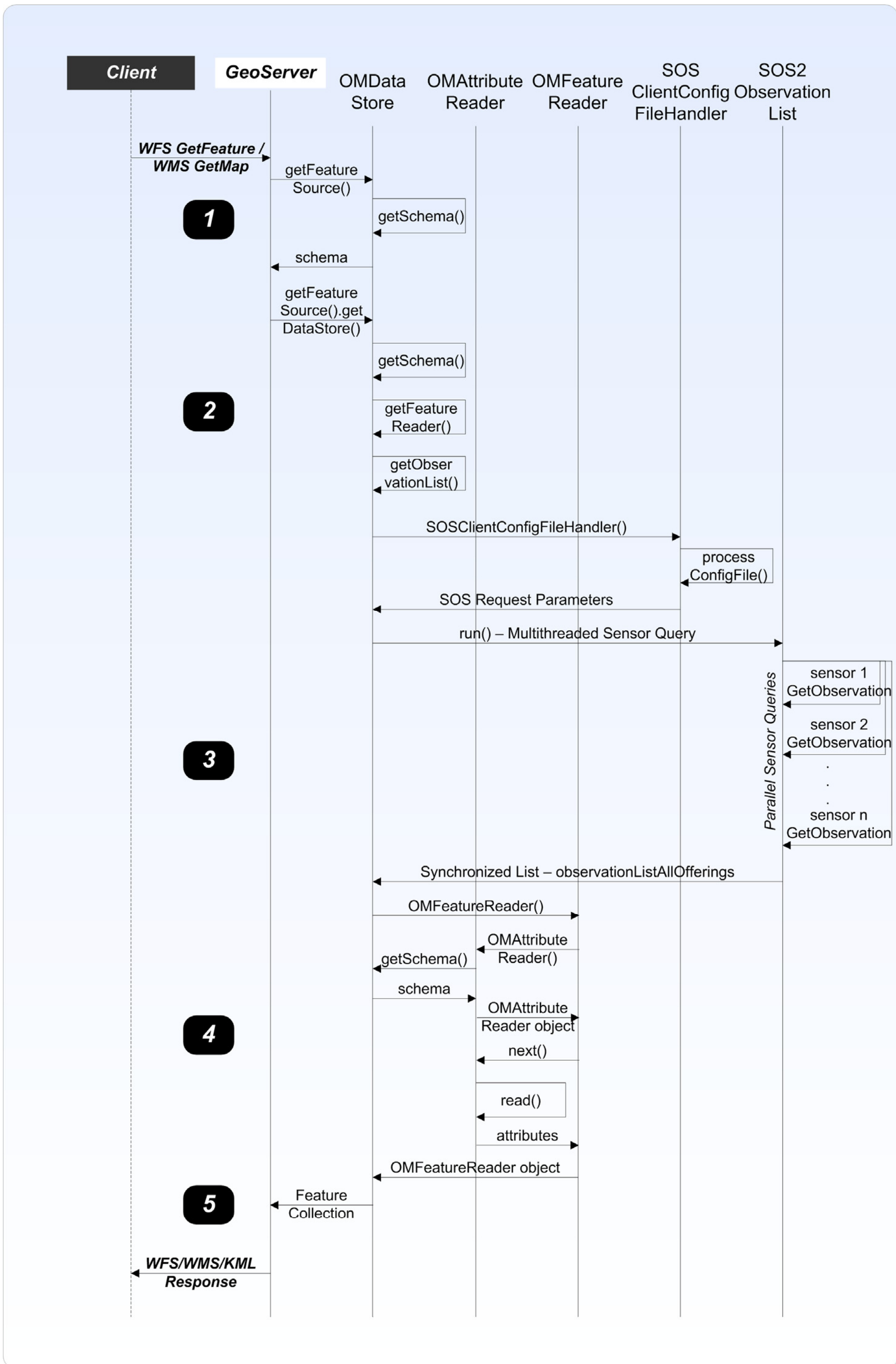

Figure 2: Sensor Fusion Service (SFS) Sequence Diagram.

5. Discussion

The presented solution has a distinct advantage versus other recent approaches. At the moment, integration of sensor data into GIS analysis systems mostly happens via the laborious interim step of a temporary physical database. This is not desirable in an automated real-time workflow as the database can easily become a bottleneck for large data volumes and bulky data sets. Moreover, this approach unnecessarily adds another component to the overall workflow, which can result in substantial performance loss.

Thus, an approach towards on-the-fly integration of sensor measurements was chosen for the Sensor Fusion Service. A substantial advantage of the GeoServer plug-in is manifested in its high flexibility and its low external dependence due to open access to GeoServer source code.

Furthermore, the compliance of the server with OGC data provision standards (WFS and WMS) offers a broad spectrum of different response formats. This allows for integration of output data not only in specialised GIS products (e.g. using OGC GML or KML), but as well outside the GIS domain by using custom formats such as Scalable Vector Graphics (SVG) or Portable Document Format (PDF). Apart from that, output data can be easily integrated into mapping Application Programming Interfaces (API) such as Google Maps or Bing Maps by using WFS output in GeoRSS or geoJSON formats.

Another benefit of this approach is that no use-case motivated and thus functionally limited web application has to be implemented for sensor data access. In fact, live measurements can be queried via well-established OGC interfaces that allow for simple integration into the geo-data visualisation and analysis portals.

6. Conclusion

This paper presents an on-the-fly sensor fusion approach to combine measurements from a variety of heterogeneous geo-sensor networks in real time. The approach largely builds on open standards such as the OGC Sensor Observartion Service (SOS), the OGC Web Feature Service (WFS) and the OGC Web Map Service (WMS).

The proposed solution shows several advantages over previous sensor fusion mechanisms such as simple integrability into existing workflows and expert tools, the elimination of bottlenecks through omitting interim databases for data storage, compliance with a number of open (geospatial) standards, and the use of open-source technologies for the realisation of the system.

At this stage, the Sensor Fusion Service is able to harmonise measurements from several geo-sensor networks by correlating basic parameters such as spatial reference systems, units of measure, time zones etc. Further research, which is currently being kicked off, aims to investigate the integration of more sophisticated fusion techniques, the combination with quality assurance mechanisms and the establishment of a feedback loop between the physical and the digital realms by using sensors along with actuators to induce changes in the sensed environment.

References

[1] Resch, B., Mittlboeck, M., Girardin, F., Britter, R. and Ratti, C. (2009) Live Geography - Embedded Sensing for Standardised Urban Environmental Monitoring. International Journal on Advances in Systems and Measurements, 2(2&3), ISSN 1942-261x, pp. 156-167.

[2] Ricker, J. (2005) Sensor Fusion: Theory and Application. Technology Paper, http://www.eetimes.com, August 2005. (21 January 2012)

[3] Simonis, I. (2008) OGC Sensor Web Enablement Architecture. OGC Best Practice Paper, Document Number 06-021r3, Version 1.0.0, http://www.opengeospatial.org, 01 July 2008. (20 January 2012)

[4] Ziegler, P. and Dittrich, K. (2004) Three Decades of Data Integration – All Problems Solved?. 1st International IFIP Conference on Semantics of a Networked World, University of Zurich, 2004.

[5] Rittman, M. (2008) An Introduction to Real-Time Data Integration. http://www.oracle.com, 2008. (22 August 2011)

[6] Sybase Inc. (2010) Real-Time Events Data Integration Software. http://www.sybase.com, 2010. (22 March 2011)

[7] Rahm, E., Thor, A. and Aumueller D. (2007) Dynamic Fusion of Web Data. XSym 2007, pp. 14-16, Vienna, Austria.

[8] Harrie, L. (2004) Using Simultaneous Graphic Generalisation in a System for Real-Time Maps. Papers of the ICA Workshop on Generalisation and Multiple Representation, Leicester, 20-21August 2004.

[9] Lehto, L. and Sarjakoski, L.T. (2005) Real-time Generalisation of XML-encoded Spatial Data for the Web and Mobile Devices. International Journal of Geographical Information Science, 19(8-9), pp. 957-973,

[10] Sarjakoski, T., Sester, M., Illert, A., Rystedt, B., Nissen, F. and Ruotsalainen, R. (2004) Geospatial Info-mobility Service by Real-time Data-integration and Generalisation. http://gimodig.fgi.fi, 08 November 2004. (22 September 2011)

[11] Srivastava, M., Hansen, M., Burke, J., Parker, A., Reddy, S., Saurabh, G., Allman, M., Paxson, V. and Estrin D. (2006) Wireless Urban Sensing Systems. Technical Report 65, Center for Embedded Network Sensing, UCLA, April 2006.

[12] Balazinska, M., Deshpande, A., Franklin, M.J., Gibbons, P.B., Gray, J., Hansen, M., Liebhold, M., Nath, S., Szalay, A. and Tao, V. (2007) Data Management in the Worldwide Sensor Web. IEEE Pervasive Computing, 6(2), pp. 30-40, Apr-Jun 2007.

[13] Riva, O. and Borcea, C. (2007) The Urbanet Revolution: Sensor Power to the People!. IEEE Pervasive Computing, 6(2), pp. 41-49, Apr-Jun 2007.

[14] Pan, A. and Viña, A. (2004) An Alternative Architecture for Financial Data Integration. Communications of the ACM (CACM), 47(5), pp. 37-40, May 2004.

[15] Choi, J., Yoo, S., Park, H., and Chun, J. (2005) Seamless Real-time Clinical Data Integration for Mobile Clinical Information System. In: Proceedings of the 21st International Conference on Data Engineering Workshops (ICDEW), Tokyo, Japan, 05-08 April, 2005.

[16] Yoo S., Kim B., Park H., Choi J. and Chun, J. (2003) Realization of Real-time Clinical Data Integration Using Advanced Database Technology. In: Proceedings of AMIA 2003 Annual Symposium, pp. 738-742.

[17] Kansal, A., Nath, S., Liu, J. and Zhao, F. (2007) SenseWeb: An Infrastructure for Shared Sensing. IEEE Multimedia, 14(4), pp. 8-13, October-December 2007.

[18] OpenGeo (2012) Welcome – GeoServer. http://geoserver.org, 17 January 2012. (18 January 2012)