On using a Sensor Observation Service as an INSPIRE-compliant download service

de Ville de Goyet N. [1], Fettweis M.[2] , Vandenberghe T. [1]

[1]Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment, Belgian Marine Data Centre (BMDC), Belgium (bmdc@naturalsciences.be)
[1]Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment, Suspended Matter and Seabed Monitoring (SUMO), Belgium

Due to the use of **sensors**, the volume of scientific data produced every day has become massive, so there is a strong need to organize them and to set up a data infrastructure for their efficient management. **Open access**, **FAIRness** (Findable, Accessible, Interoperable and Re-usable) and **INSPIRE-compliance** are increasingly becoming the norm for (environmental) data management. In order to achieve complex aspirations such as FAIRness and INSPIRE-compliance a simple idea might help: "Collect Once, Use many times". Data collected today must be stored, documented and published in order to increase their knowledge extraction and to allow for an efficient re-use in the future. With the multiplication of sensor deployments in monitoring programs, the new challenge is to publish time-series efficiently using state-of-the-art technologies.

The MOMO project (MOnitoring and Modeling of the cohesive sediment transport and the evaluation of the effects on the marine ecosystem resulting from dredging and dumping operations) has been deploying tripod platforms consisting of ADP, OBS, LISST and ADV sensor packages in the Belgian Part of the North Sea for the last two decades. These deployments generated considerable amounts of valuable data that are currently stored as csv-formatted text files that are not interoperable, both in terms of content and access protocol. The full potential of those data is currently safeguarded only thanks to the knowledge and expertise of the scientists involved in the project. The risk of knowledge loss is therefore significant.

The **objective** of this work is to explore the merit of implementing an Open Geospatial Consortium Sensor Observation Service (OGC SOS) for **the publication of FAIR INSPIRE-compliant long time-series data**.

FAIRness and INSPIRE-compliance

To achieve an INSPIRE-compliance a complete environment must be created with metadata, services and data. Most of the work is done by now as the official RBINS metadata catalog is online (**https://metadata.naturalsciences.be**) with INSPIRE-compliant metadata and services. (Meta)Data are now discoverable but the Interoperability still remains a challenge.

The Sensor Observation Service has been recognized as an interoperable INSPIRE-compliant download service and the 52°North implementation out-of-the-box includes the INSPIRE O&M Specialized Observation (OMSO) profile. To deliver interoperable data only three steps are needed in our case: create the database, populate it and connect the SOS client to the metadata file. Pre-defined datasets are then shared using the SOS operation GetObservation in the metadata file. This approach achieves FAIRness and INSPIRE-compliance (see dotted-line box in Fig. 1).

Achieving full INSPIRE-compliance requires also creating a view service (which shows a map). We have exposed a view on the Sensor Observation Service's database as a WMS feature using GeoServer **(https://spatial.naturalsciences.be**) this might prove an effective way to publish sensor data in the INSPIRE context (see Figure 1).

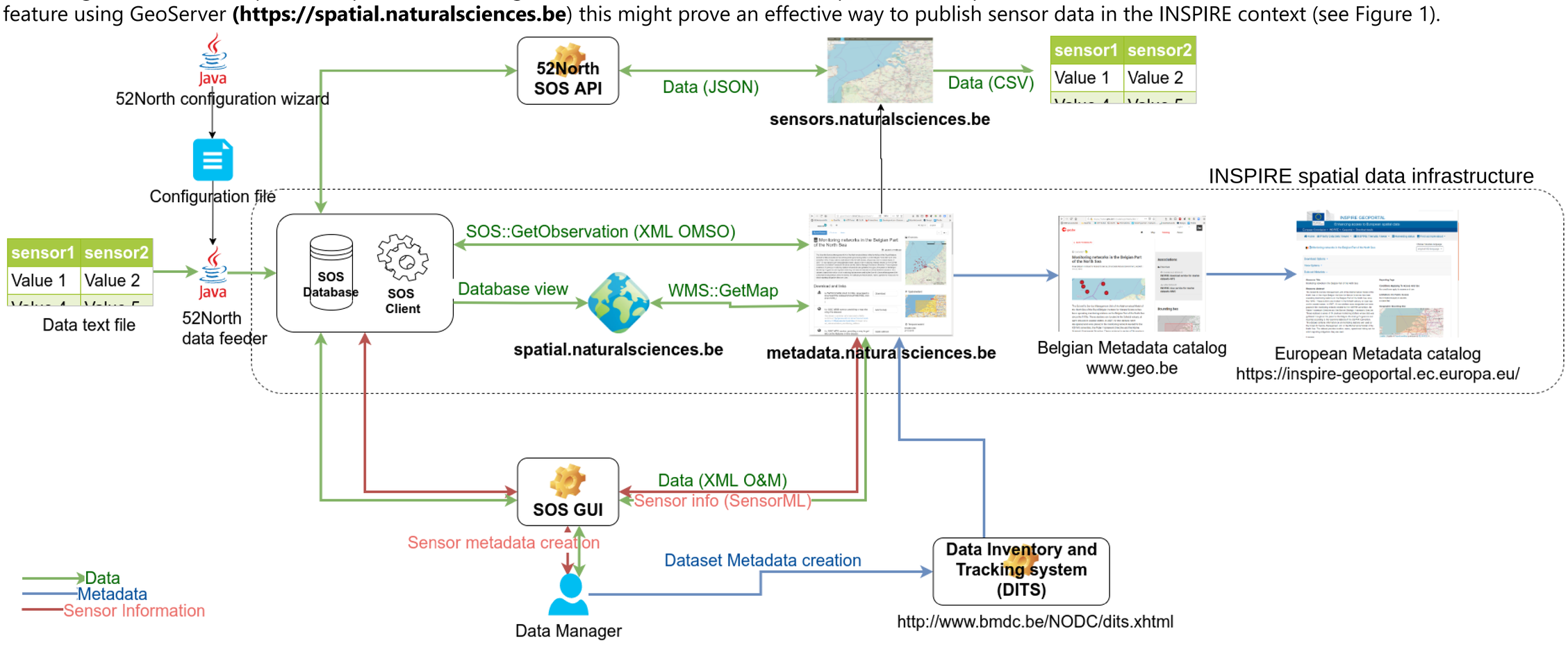

Figure 1: Schematic data flow for the publication of time series data through SOS. The dotted-line box indicates the mandatory elements for INSPIRE-compliance. Additionally the 52North SOS implementations comes with a web-viewer built upon the SOS API (sensors.naturalsciences.be) where time-series can be visualized and downloaded as csv files.

Performances

The publication of time-series automatically comes with the burden of ensuring download service performance. The amount of data generated each year in this project is massive and providing those data in INSPIRE-compliant OMSO standard is challenging. This is due in part because interoperable data is inherently verbose.

A first attempt was to import all the project data in the PostgreSQL database (+150e6 values from four sensors located at multiple locations). Metadata files were created for each dataset and a GetObservation request was made to retrieve the dataset data. However, each time-series contain millions of points and the big disadvantage of heavy XML encoding quickly becomes obvious. Download requests ends up in time-out errors.

The second attempt was to import only the "aggregated" measurement (i.e. measurements that don't need to be associated to other measurements to have a value). The main time-series in the monitoring programme such as temperature, salinity or sediment concentration are published via the SOS service and multi-parameters sensors (more than a hundred for Acoustic Doppler Profiler sensor) are left aside. The size of the database remains much smaller and the service is able to react much more quickly.

Conclusions

The promotion of FAIRness, INSPIRE-compliance or open data rises the hope to share more environmental data in a way that will help increase the global understanding of our environment. Public data has to be shared with the widest public and private community and this, in order to be efficient, needs a high level of interoperability. INSPIRE is the European Directive that pushes the implementation of interoperable spatial data infrastructure for the Members States.
Open-source initiatives are complete and mature enough for the set-up of a complete INSPIRE-compliant data infrastructure (e.g. geonetwork/geoserver/SOS client) even if there are still some challenges for the seamless integration and parametrisation of the different components. Once the system is up and running, there is a strong temptation to publish everything. However, the performance of the protocols does not allow to massively publish sensors data as open data but forces the choice of smaller sub-sample datasets.
The optimal solution, at this time, is to choose for a mixed approach with SOS for standalone time-series and a files-based system for sensors that generate too many data. As suggested in the INSPIRE specification an out-of-the-band gml encoding could be used to describe the data in text files but the interoperability of this method still has to be demonstrated.