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

**Nicolas de Ville de Goyet,** Royal Belgian Institute of Natural Sciences (Belgium),
ndeville@naturalsciences.be
**Michael Fettweis**, Royal Belgian Institute of Natural Sciences (Belgium),
mfettweis@naturalsciences.be
**Thomas Vandenberghe,** Royal Belgian Institute of Natural Sciences (Belgium),
tvandenberghe@naturalsciences.be

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 programmes, the new challenge is to publish time-series efficiently using state-of-the-art technologies.

The MOMO project (MOnitoring and MOdelling 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 ADCP and ADC 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 consequences 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 time-series data. The INSPIRE directive must be applied to data by the end of 2020 so the question is: “can a data provider be INSPIRE-compliant using SOS without complex and heavy developments?”

The 52°North SOS implementation was chosen in this project, partly because they provide a complete open-source solution (database creation, SOS client and SOS viewer).

**FAIRness and INSPIRE-compliance**

To achieve a FAIR (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 Findable, Accessible and Re-usable but the Interoperability still remains a challenge.

The Sensor Observation Service has been recognized as an 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 two steps are needed: create the database (schema provided by the SOS client) and populate it. Pre-defined datasets are then shared using the GetObservation SOS operation in the metadata file. This approach achieves FAIRness and INSPIRE-compliance.

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); this might prove an effective way to publish sensor data in the INSPIRE context (see Figure 1).

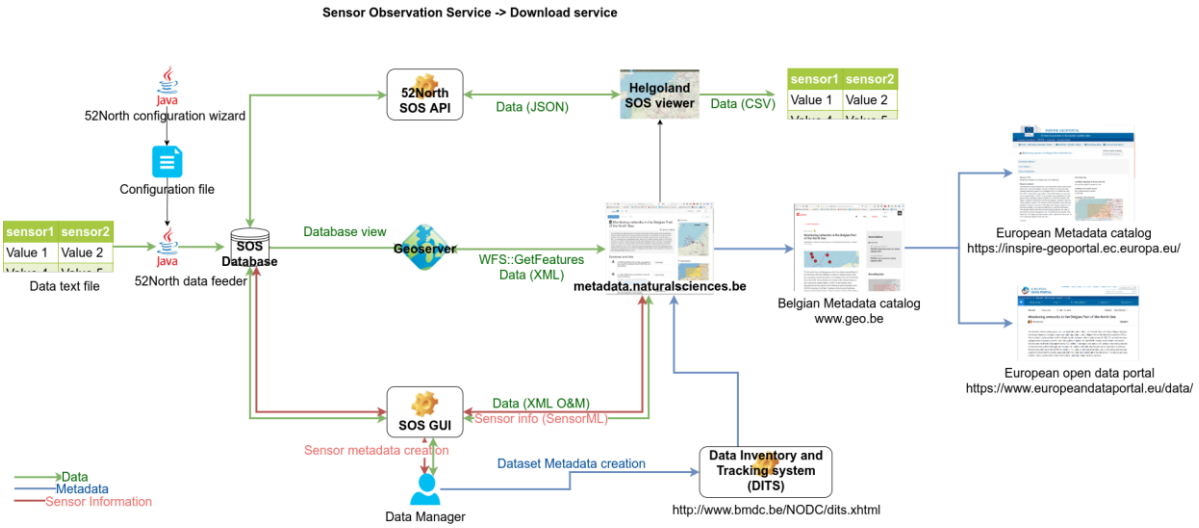


Figure 1: INSPIRE-compliant data publication workflow from text files to metadata publication.

**Performance**

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 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. We are of the opinion that the OMSO format is not optimal for time series publication as many information elements are redundant between observations. This leads to download requests ending up in time-out errors. If one wants to publish INSPIRE time series the use of O&M time series format should be studied but it is not available in the 52°North implementation.

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. 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 time-series. An out-of-the-band gml encoding could be used to describe the data in the files but the interoperability of this method still has to be demonstrated.