

# IC\_14 SOS & SSN ontology based data acquisition and NRT technical innovations

## 1. Background

### 1.1 Short description

**Data Acquisition** services and in particular the preparation of data transfer (ENVRI RM: **prepare data transfer**) prior to data transmission are not yet sufficiently standardized. This hinders efficient, multi RI (Research Infrastructure) data processing routines such as **Data Quality checking**. This use case intends to promote standardization and move the standardization level close to the sensor. Objectives include:

- Standardized data transmission using OGC SWE Transactional SOS (Sensor Observation Service) as priority standard as well as using the Semantic Sensor Network (SSN) ontology. Both will be implemented and tested.
- Generic quality control (QC) routines suitable for multiple RIs (e.g. EMSO, EuroARGO, ANAEE, etc.) will be defined and implemented at own and/or EGI platforms.
- QC routines will be used to process these standardized data transmission streams to enable Near Real Time (NRT) quality control routines on raw data.

### 1.2 Contact

<b>Background</b>	<b>Contact Person</b>	<b>Organization</b>	<b>Contact email</b>
RI-ICT RI-ICT RI-ICT	Robert Huber Andree Behnken Markus Stocker	UniHB, PANGAEA, EMSO UniHB, PANGAEA, FixO3 UniHB,PANGAEA	<a href="mailto:rhuber@uni-bremen.de">rhuber@uni-bremen.de</a> <a href="mailto:abehnken@marum.de">abehnken@marum.de</a> <a href="mailto:stocker@marum.de">stocker@marum.de</a>
RI-ICT	Thierry Carval	Ifremer, Euro-Argo	<a href="mailto:thierry.Carval@ifremer.fr">thierry.Carval@ifremer.fr</a>
RI-ICT	Olivier Gilbert	EPOS	<a href="mailto:olivier.gilbert@univ-grenoble-alpes.fr">olivier.gilbert@univ-grenoble-alpes.fr</a>
RI-ICT	Mauro Mazzola	SIOS	<a href="mailto:m.mazzola@isac.cnr.it">m.mazzola@isac.cnr.it</a>
RI-ICT	Federico Carenuto	ANAEE	<a href="mailto:carotenuto@ibimet.cnr.it">carotenuto@ibimet.cnr.it</a>
RI-ICT	Alessandro Zaldei	ANAEE	<a href="mailto:a.zaldei@ibimet.cnr.it">a.zaldei@ibimet.cnr.it</a>
RI-ICT	Alessandro Matese	ANAEE	<a href="mailto:a.matese@ibimet.cnr.it">a.matese@ibimet.cnr.it</a>
RI-ICT	Simona Scollo	EPOS	<a href="mailto:simone.scollo@ingv.it">simone.scollo@ingv.it</a>
RI-ICT	Jean-Francois Rolin	EMSO	<a href="mailto:jean.francois.rolin@ifremer.fr">jean.francois.rolin@ifremer.fr</a>
RI-ICT	Mickael Langlais	EPOS	<a href="mailto:mickael.langlais@univ-grenoble-alpes.fr">mickael.langlais@univ-grenoble-alpes.fr</a>
RI-ICT	Fadi Obeid	Lab STICC	<a href="mailto:fadi.obeid@ensta-bretagne.org">fadi.obeid@ensta-bretagne.org</a>
RI-ICT	Angelo Viola	SIOS	<a href="mailto:angelo.viola@artov.isac.cnr.it">angelo.viola@artov.isac.cnr.it</a>
RI-ICT	Thomas Loubrieu	SeaDataNet	<a href="mailto:Thomas.Loubrieu@ifremer.fr">Thomas.Loubrieu@ifremer.fr</a>
RI-ICT	Jean-Francois Rolin	IFREMER	<a href="mailto:Jean.Francois.Rolin@ifremer.fr">Jean.Francois.Rolin@ifremer.fr</a>

### 1.3 Use case type

The use case will be an implementation case.

### 1.4 Scientific domain and communities

#### Scientific domain

All

#### Communities

Data Acquisition, Data Service Provision

#### Behaviours

Relevant community behaviours: Instrument Configuration, Data Collection, Data Quality Checking, Semantic Harmonization

Relevant community roles: Sensor, Sensor Network, Measurement Model Designer, Data Acquisition Subsystem, Data Curator, Semantic Curator

## 2. Detailed description

### Objective and Impact

The use case will move the standardization level close to the sensors of RIs, thus allow the implementation of common, generic data processing routines such as NRT QC .

'Data Transmission' at the sensor as well as platform level (Fig.1) largely depends on community specific needs and habits or simply on manufacturer specifications. Both result in proprietary or niche formats and protocols that require data to subsequently be processed by data transformation services before they can be delivered in a standardized format (Fig. 1). The ENVRIPLUS objective in WP1 is to promote sensor web enablement strategies for the various context of RIs it is then of high importance to connect the choices in sensor interface and the Quality control procedures.

The objective of this use case is to contribute to the harmonization of data transmission formats and protocols.

The use case will test two data transmission formats and protocols, namely the OGC Sensor Web Enablement (SWE) Sensor Observation Service (SOS) and the Semantic Sensor Network (SSN) ontology in combination with RDF Streams<sup>[2]</sup>. Appropriate data compression (EXI) will be tested to ensure resource friendly data transmission.

Furthermore, it will implement generic quality control procedures such as those defined within WP 3.3, and make use of the standardized data transmission formats to perform generic, cross-RI NRT QC routines, tag the controlled data with appropriate data quality flags, again using the standard formats mentioned above.

## Challenges

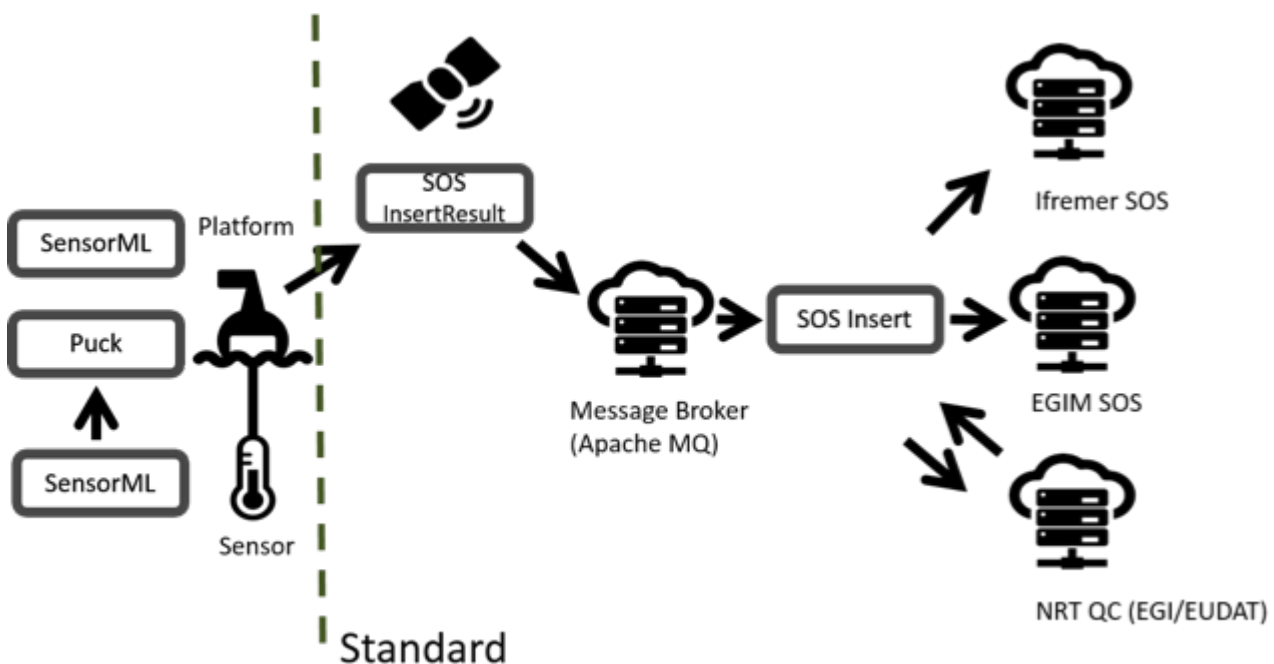
- Provide data as well as metadata on sensors and data in a standardized way
- Enable data transmission which is standardized, sufficiently described with metadata as well as resource friendly.
- Provide generic data quality routines that are relevant to most RIs

## Detailed scenarios

The use case will test two scenarios. One will be based on the Sensor Web Enablement (SWE) suite of standards (e.g. SOS) while the other will use the Semantic Sensor Network (SSN) ontology.

The Sensor Web Enablement (SWE) approach

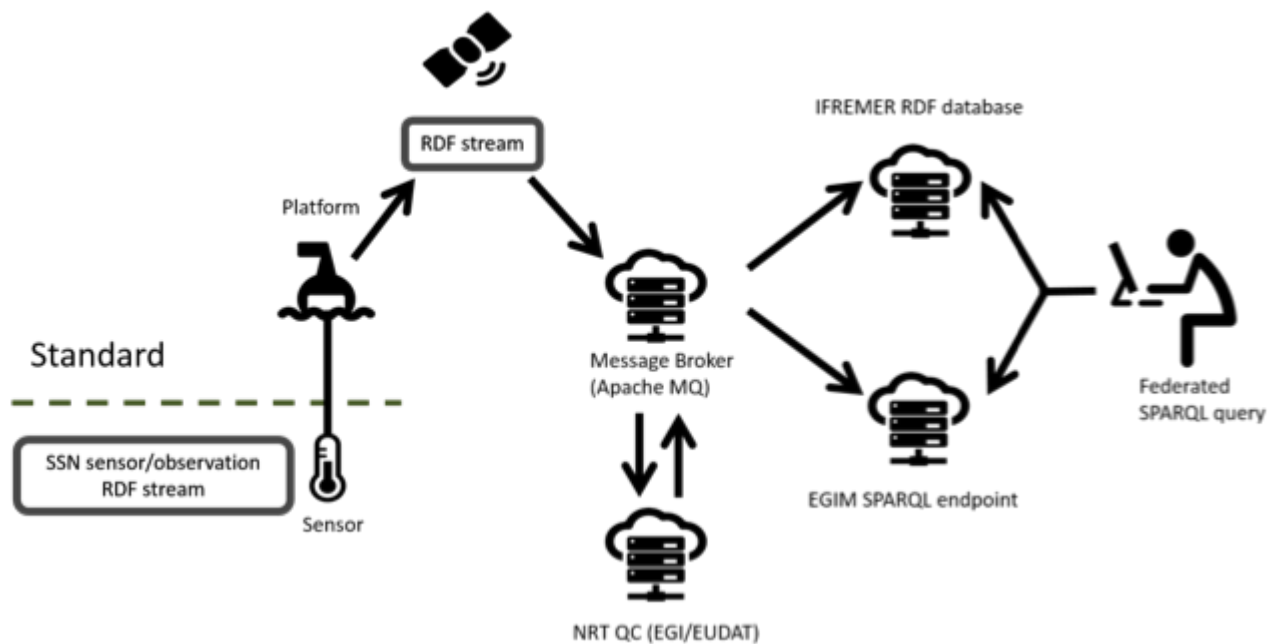
The idea is to use transactional SOS requests to transmit data. The sensor/platform will transmit data via transactional SOS InsertObservation commands. An (optional) message broker will forward these commands to the service endpoint at e.g. EGI and at the same time send the data to the RI data processing center. At EGI and/or the RI data staging endpoint NRT QC routines will take place.



Tasks:

- Define Transactional SOS requests and templates
  - Templates
  - InsertObservation request
- Implement Transactional SOS requests (dummy and/or at use case platform)
- Implement compression service (e.g. EXI)
- Install SOS server (or adopt PANGAEA's SOS) which supports transactional SOS to receive and process InsertObservation requests.

The Semantic Sensor Network (SSN) ontology approach



## Required services:

We have described an ENVIplus Implementation Case that aims at embedding standards for the encoding and format of observation data into sensing devices. Of specific focus are standards by the OGC - in particular SensorML, Observation & Measurements, and Sensor Observation Service - and recommendations by the W3C - specifically the Semantic Sensor Network ontology. Embedding such standards into sensing devices enables the acquisition of observation data from sensing devices natively encoded and formatted following these standards. This will reduce the number of translations required during data acquisition. Given standardized streams of observation data, the Implementation Case investigates the execution of generic data processing routines on data streams. Of interest are routines for near real-time quality control (NRT QC).

We think that Apache Storm [1] could support the implementation of the proposed case. Apache Storm is a distributed real-time computation system. It specializes on reliable processing of data streams and is designed to support real-time analytics and continuous computation, among other use cases. Central to Apache Storm is the notion of Storm topology. A topology consumes streams of data and processes streams in arbitrarily complex ways. It thus models the logic for a real-time application. A topology is a directed acyclic graph. Nodes are either spouts or bolts. Vertices are streams. A stream is an unbounded sequence of tuples. Tuples are data packages. A spout is a source of streams in a topology. Bolts perform computations (processing) on tuples.

We intend to investigate the application of Apache Storm to the described Implementation Case. Specifically, the idea is to model the data acquisition and NRT QC computations as a Storm topology. Sensing devices may be modelled as Storm spouts, i.e. as sources of streams. Streams model the transmission of sensor data in the topology. The data - here encoded following the OGC and/or W3C standards - are modelled as tuples. Finally, any computational node is modelled as a bolt of the topology. Of particular interest are computational nodes that execute a routine for NRT QC, such as outlier detection. Being modelled as bolts of a Storm topology, such outlier detection operates as a continuous computation task on the tuples of streams, as specified by the topology.

We also intend to partner with EGI which could serve as platform for the deployment of Storm topologies on a distributed computer network.

[1] <http://storm.apache.org/>

## Tasks:

- Define SSN templates for sensor data and metadata
- Implement the representation of SSN-conformant sensor metadata and data in sensors (dummy and/or at use case platform, in any case consistent with the sensor networks addressed in Task 1.4)
- Implement transmission of RDF stream data (stream of triples) to a triple store
- Provided access to SSN-conformant sensor metadata and data via a SPARQL endpoint

## Near real-time quality control

Tasks:

- Define some basic NRT QC procedures (based on Zandvoort discussion, the WP3.3 evaluation (and e.g. <https://ioos.noaa.gov/project/qartod/>))
- Implement a SOS InsertObservation unpacking and **transformation service**
- Implement some basic **QC services** which use transformed SOS data as input
- Evaluate EGI services for applicability (e.g. virtual machines)
- Deploy both the transformation service and the QC service at MARUM and/or EGI
- Testing and analysis
- Based on the results from implementing QC services on SOS data, prototype examples for NRT QC on RDF streams

## Technical status and requirements

The use case involves EGI and requires, e.g., the use of a scalable data processing environment via virtual machines.

## Implementation plan and timetable

1. Month 5: Implementation of SOS based data transmission
2. Month 10: Implementation of SSN based data transmission
3. Month 12: Generic data object model (e.g. based on SSN)
4. Month 15: VM for data brokering and data quality control routines
5. Month 18: Deployment of NRT quality control routines

## Expected output and evaluation of output

A generic NRT QC service capable of accepting standardized SOS data or SSN RDF data streams will be ready.

[2] <https://www.w3.org/community/rsp/>