Project contract no. 036851

ESONET

European Seas Observatory Network

Instrument: **Network of Excellence (NoE)**
Thematic Priority: **1.1.6.3 – Climate Change and Ecosystems**
Sub Priority: **III – Global Change and Ecosystems**

D70 - Updated Data Management Plan

Due date of deliverable: month 36
Actual submission date: month 37

Start date of project: **March 2007** Duration: **48 months**

Organisation name of lead contractor for this deliverable: Participant # 24B, UPC,Spain
Lead authors for this deliverable: Mike van der Schaar, Michel André, Robert Huber

Revision [1, 06.04.2010]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	**Public**	**X**
PP	**Restricted to other programme participants (including the Commission Services**	
RE	**Restricted to a group specified by the consortium (including the Commission Services)**	
CO	**Confidential, only for members of the consortium (including the Commission Services)**	

Contents

1. Introduction

This document is a continuation of the ESONET Data and Information Management Plan (Project Deliverable D9, Data Management Plan, Diepenbroek et al, 2007) which gave a broad overview of existing data management policies, their importance and how these relate to the ESONET. This document aims to give a more detailed overview of how a specific data management policy can be implemented at a site. As a test case, interoperability between the LIDO DEMO mission and the WDC-MARE archiving centre is tested. It is anticipated that the ANTARES platform will follow the guidelines in this report once the ALBATROSS DEMO mission becomes operational (this is not expected before November). As many research fields already have their own established ways of handling, storing and documenting data, the intention here is not to enforce a specific standard for the entire ESONET. Its aim is rather to function as a roadmap for these cases where a data management plan has not yet been established and where it is not clear what standard to follow.

There are different standards available for handling (meta-)data and data distribution. Considering that the SensorML specification (currently part of the Sensor Web Enablement (SWE) initiative defined by the Open Geospatial Consortium (OGC)) was already extensively used for the description of hardware, it seemed most obvious to continue with this line and look at the other specifications that are available under the SWE. To minimise the number of protocols that need to be supported this document focuses on (and recommends the use of) SWE formats and especially the SensorML, Observation & Measurement and Sensor Observation Service parts of the protocol. The SWE standards unfortunately include a plethora of namespaces that can be used to encode the XML files. Namespaces that will be of interest in these recommendations will be **sml**, **gml** and **swe**, with the addition of **ows** for error handling (http://schemas.opengis.net/). For the definition of data types the W3C specification was used (http://www.w3.org/TR/xmlschema-2).

Data distribution

A Sensor Observation Service is being developed within the ESONET by Ifremer for the distribution of data. Its current development is tied closely to the Coriolis database and it might not be immediately deployable by third parties with a different database structure. An alternative SOS server that is easier to set-up and test with is the Oostethys PERL server (http://www.oostethys.org/downloads). It offers support for template-based queries for the GetCapabilities (retrieval of information about the platform, the sensors and available data), DescribeSensor (retrieval of information on a particular sensor or procedure) and GetObservation (actual measurements) requests. It is expected that this server will provide an easy test platform for sites that consider using SOS for data distribution, which can then later be replaced by the more complete implementation from the Ifremer.

Meta-data description

Following the current use of SensorML in the ESONET to describe sensors, the same specification is suggested for the description of meta-data. Meta-data should be defined for each data set that is distributed, where a data set is a collection of data gathered over a specific time interval. The meta-data may change between data sets from the same site. The description should include at least:

- sensor information (possibly through a link to a specific SensorML file)
- procedure information (experiment set-up)

- algorithms (some measurements might the result of operations performed on the raw data)
- data descriptions (data types, bounds)
- institute/citation information
- clock information

Originally, it was thought that the O&M procedure element could be used for the description of procedures or algorithms, but it is not allowed to include O&M inside SensorML[RH1]. As detailed below, SensorML does provide a way to describe 'non-physical or pure processes' (ProcessModel/ProcessMethod), but this was found to be somewhat complex and as an alternative it is suggested to put information inside documentation elements.

For archiving and exchange purposes, the ISO19136 XML implementation of the ISO19115 metadata schema to describe, validate, and exchange geogreferenced data should be used. The choice of ISO 19139 is supported and driven by INSPIRE Implementing Rules and is also used by other large project such as SeaDataNet or IODP. Thus it is best suited to represent ESONET metadata.

Data encoding

The data can be encoded using the O&M specification, following its definition in the meta-data file. The meta-data itself can be included in the file with the data (in SWE, O&M allows the inclusion of SensorML but not vice versa), or a link could be included to the relevant information. In the latter case it is assumed that the meta-data does not change much (if at all) between data sets.

Special consideration is given here to hydrophone data. Many examples and description of the SWE standards assume data that is sparsely sampled, a few measurements per minute, hour or day. In those cases, it is feasible to encapsulate the data in an XML file and distribute it as text. A real-time data stream that provides access to this data will have very low bandwidth requirements. This situation already changes when digital camera images are considered, but changes completely with the data flow of one or more hydrophones used for cetacean detection. Typically, data will be sampled at 96000 Hz (preferably at a double rate when dolphin sonar is to be recorded) and encoded in at least 16 bits. The required bandwidth for a single hydrophone then becomes 190 kB/s. Clearly, distribution of the data in text form is no longer useful and the alternative is to distribute the data in binary packages with a description of the data format in the meta-data (following SWE standards).

Considering the problem of distributing binary files, endianness for example, one standard that can be used is the network Common Data Form (NetCDF, http://www.unidata.ucar.edu/software/netcdf/). This format provides a machine independent description of the data allowing easy exchange. However, the format stands on its own, it does not use XML descriptions of the data format and it does not fit in well with the OGC SWE standards. It would require extra libraries compiled into the code, in addition to those needed to parse XML files for SWE standards, complicating code and procedures. For this reason it is not considered here for distribution of binary data. Binary distribution is not considered at all in this document (WDC-MARE will not yet store raw acoustic data, but will initially focus on derived data).

As a last remark, the complexity of the SWE standards, especially the inclusion and nesting of many different XML name spaces, but also the intensely verbose description of its elements, makes it difficult to manually construct these files. It can be strongly advised to use an XML editor such as Oxygen to create and validate templates. The ESONET will provide SensorML templates for sensors, but is not currently doing the same for (meta-)data descriptions. This may block standardisation as it cannot be expected that everyone goes through the tedious process of figuring out how to specify things in the verbose OGC standards.

Appendix I contains the results of a small questionnaire sent to the DEMO projects in the ESONET concerning their plans on data management. If there is one conclusion that can be drawn from this, it is that data management is generally an afterthought and not a main concern. With that in mind, the recommendations have to be kept as simple as possible, with the use of as few different protocols as possible.

Appendix II contains a summary of a data management meeting that was held in Vilanova i la Geltrú under the framework of an exchange of personnel. The recommendations in this document are partly based on the results of this meeting.

Appendix III contains examples of (validated) XML files used for the connection between LIDO and the WDC-MARE that could be used as templates by other projects.

2. Overview of data acquisition process

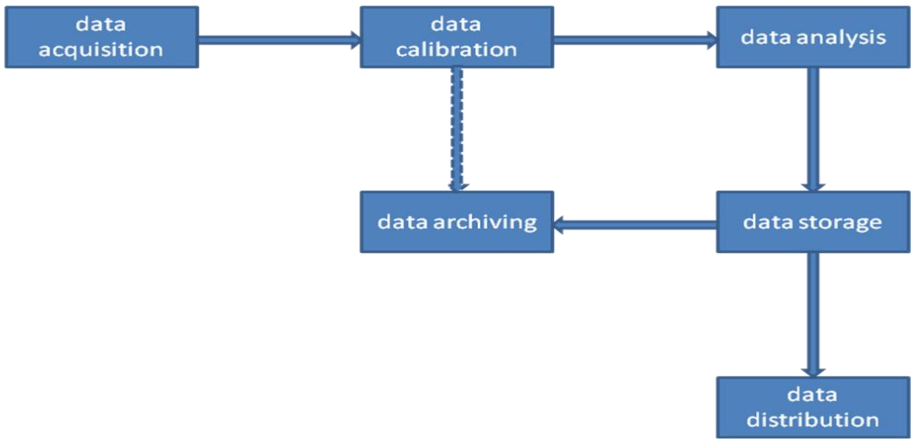

Figure 1. Stages in the data acquisition process

Several steps can usually be identified between the acquisition and distribution of data. All these steps need to be carefully documented to allow third parties to understand and use the data that is distributed. The different steps are shown in Figure 1. Stages in the data acquisition process, with arrows corresponding to how data may flow through the process. However, not all steps may always be present or followed in this order.

- data acquisition

    This is the step where data is acquired by hardware, which can be a single sensor or a system with different components (e.g. a hydrophone with a pregain amplifier, conditioning amplifier and digitizer). Within the ESONET, the description of the hardware is intended to be made available using SensorML, which should include all information on what is done to the data at the hardware level. From the data management point of view, this step will not need any further description, other then that a link to the hardware SensorML description should be included in the meta-data.

- data calibration

  Some calibration of the data may be performed at the hardware level (signal conditioner), but most often this will be a separate step after the acquisition. Based on discussions during various workshops, it will be recommended not to distribute data without calibration. The main reason for this is that it is easy to misinterpret the meaning of the data without calibration, and that overall, the raw data itself can be considered useless for scientific purposes. To allow recalibration of data at a later point, it can be useful to store or distribute the raw data together with required calibration information in the meta-data, instead of applying calibration directly. The information needed for calibration should either be included in the SensorML description of the hardware or in the meta-data description of the data set.

- data analysis

  At least two types of analysis can be identified at this step. First there can be a quality assessment of the data. Human supervision on the data is often not possible when data is distributed in real-time, but it is still preferred to have some qualification on the quality of the data to see if it should be distributed at all, or to allow a decision to be made on the usefulness of downloading it. An automatic quality control will be basic and its 'quality' will depend on the a-priori knowledge about the signal that is acquired. At the very least, it should be able to detect if there is any signal at all (there might just be electrical noise when some part of the acquisition system became disconnected) or if there is complete saturation. Other then that, knowledge of the signal might allow more precise bounds on the signal level (or measurement outcome in general) to decide what can be considered a reasonable measurement or simply an error.

  Apart from quality assessment, other analysis may be performed on the data, which outcome can also be considered a measurement by itself and can be distributed. This analysis can include the detection and classification of specific events, for example in the LIDO case, the detection of dolphins from hydrophone data. In order to interpret or understand the results of this additional analysis on the data, it is important to document the process or algorithms that were used for their computation in the meta-data.

  The documentation of the analysis that is done on the data will be described further below.

- data storage

  In the presented overview of data management steps, data storage is considered short time storage. Additionally, based on the outcome of the quality assessment, not all data may be stored; some data may be immediately discarded if it was found to be useless.

  In many cases data will be stored temporarily on-site (where it was acquired) and later moved to a storage facility elsewhere. Real-time distribution of this data will be through direct access from this temporal storage pool, and the storage formats may not be the same as those used for permanent archiving.

The protocols used for deciding whether or not data is kept and for how long it will be available should be documented in the description of the acquisition process, which is included in the meta-data.

- data archiving

  Data archiving consists of the long term storage of the acquired data. In some cases the data will still be directly available for distribution through archival centres (e.g. WDC-MARE), but in other cases the long term archive will simply consist of a stack of hard disks on a shelf. In the latter case, this data may not be directly available for distribution and its storage format might not be easily readable by other software over time. In some cases, data may be archived directly after acquisition, without any further processing or quality assessment.

  The procedure used for archiving data (and its availability) should be included in a data archiving agreement between the data provider and the archive, in which the procedures are defined.

  When data is distributed in real-time it can be made available through a Sensor Observation Service. Temporary storage will only make available a limited amount of most recent data, as specified in the meta-data, and knowledge on the currently available data set can be obtained through the GetCapabilities request. Measurements themselves are encoded in O&M and obtained through corresponding GetObservation requests (these do not necessarily have to include meta-data descriptions as the GetCapabilities/DescribeSensor request can be used for that). This set-up will be described below and was tested between LIDO and the WDC-MARE.

3. Encoding of (meta-)data

To document the data analysis and data types/formats, it will be recommended to create an XML schema that incorporates these. While SensorML includes a framework to describe a process and algorithms used for analysis (ProcessModel/ProcessMethod), this description is both very detailed and restrictive in order to allow automated reproduction of the analysis results on the raw data by third parties. This is often not required, and it will be much simpler to explain the meaning of a value (or how it was obtained) together with its definition in a schema file. The elements that allow this are <annotation> and <documentation>, which should be used together,

```
<xs:element name="Example">
  <annotation><documentation>
    Descriptive text can be put here, including references to papers, etc.
  </documentation> </annotation>
</xs:element>
```

A larger example is given for LIDO in Appendix III.A, where a schema file is presented to encode results that are obtained from the analysis of a segment of raw acoustic data. Including boundaries for all numerical values will allow e.g. the archiving centre to perform a basic quality check on the data. In the example various measurements on the data are returned in the results, together with a link to the hardware description of the site (to a SensorML file). One of the outcomes for LIDO are probabilities for the presence of specific acoustic events, or marine mammal species, in the area. These are restricted to be between 0 and 100, and to understand the quality of these automatic detections the meta-data file (schema file here) provides some details and references to papers about the algorithms that were used for the detections.

Encoding of the data will then be straightforward and its correctness can easily be verified against the schema file (example of encoded data for LIDO is provided in Appendix III.B). For distribution of the data it is recommended to use the Sensor Observation Service. Templates will be defined for the SOS requests based on the tests between the WDC-MARE and LIDO and the examples from the Oostethys perl server. Examples will be given in Appendix III.C.

4. Interoperability in practice, LIDO - WDC-MARE

Tests for the distribution of (meta-)data using the OGC SWE protocols are being prepared between LIDO and the WDC-MARE. It is expected that results of these tests can be included into an updated version of this report as part of deliverable D43 (infrastructure productive version).

5. Recommendations for ESONET platforms

Recommendations for data management will be given based on the testing in Section 4. Very likely the recommendations will include to:

- Provide an overview of the sensors (and set-up) in SensorML (ESONET provides templates for various sensors)
- Define the process of how data is managed, it remains to be decided where this is best placed.
- Encode data in XML files when the data flow is not 'too' high
- Provide schema files for the data encodings that allow automatic validation and interpretation of the data
- Offer public distribution of the data through the Sensor Observation Service, initially using the Oostethys PERL server and templates for the main requests (as provided in Appendix III).
- Within a SOS response, provide data in O&M format or simple tab separated text

Appendix I: Data management procedures in ESONET DEMO missions

The following overview was based on a questionnaire sent in September to the demo missions within the ESONET. When there were no answers the data was collected from the project specifications.

MOMAR, taken from D12 annex :

Data management procedures will be fully compatible with international recommendations and standards in order to improve interoperability with other systems and to ease comparison with other datasets: ISO standards for metadata, COI/WMO standards for quality flag scale.

SISMER will collect, flag and archive the data (in real time and after the recovery). Data will be made available online according to ESONET data policy and European directives. Data will also be forwarded to data centers involved in the ESONET project in order to be permanently archived and distributed.

MARMARA, as taken from D12 document:

WP 4 : Data integration and modeling. This work under this WP has not started yet. It will start as soon as the data from the MarNaut piezometer and flowmeters have been processed. The work will continue after the 2009 cruises.

"Note : Only WP1, WP2, WP3 and WP6 have started before September 30, 2008. Hence no management report is provided for WP4 and WP5"

AOEM, taken from project documentation :

Data transfer formats / protocols

This work-package will evaluate formats and protocols used for transfer of data streams from other existing operational oceanographic observatories for applicability to the diverse oceanographic and acoustic data sets delivered from Fram Strait moorings via intermediate acoustic links to the cabled node. The evaluation of results from WP4.2 will define data formats and protocols for pre-processing, compression, transfer and storage of oceanographic and acoustic data from Fram Strait moorings and platforms via the cabled observatory.

Outreach :

Website with synopsis of research objectives, technology development, and near real-time data feeds from different sensor suites, and blog sites during research cruises in 2010, 2011, and 2012, and access to photos and video clips, and podcasts.

MODOO, answer by Maureen Pagnani :

The MODOO data management is planned to be an extension of the methods I am currently using for EuroSITES.

The real-time data will flow via NOCS which is acting as the EuroSITES DAC (Data Assembly Centre) and will be processed into OceanSITES format. The file storage areas used for the data are all backed up automatically within the NOCS IT system. The processed data are written to the NOCS ftp site where Coriolis, which is a GDAC (Global Data Assembly Centre) for OceanSITES, accesses it and places the files on the OceanSITES ftp area. Following recent discussions at an OceanSITES data management meeting the SensorML version of metadata, which is currently being developed, will be created to provide the information required for the SOS (Sensor Observation Service) which is being developed as part of EuroSITES. I understand that Cecile Robin is already working with Christoph Waldmann to ensure that the OceanSITES SensorML outputs will mesh with the ESONET Sensor Registry requirements.

One of the crucial outputs of MODOO will be a demonstration of the EuroSITES data management system and how it will be applicable to ESONET.

LOOME, answer by Dirk de Beer :

- *Where will the data delivered by the DM be stored initially ?*

With the different partners from LOOME, who supplied the instruments

- *Is the data accessible via the internet, if so where ?*

We will publish the data after retrieval and analyses. Retrieval will be september 2010.

- *Which data center will be responsible for the long-term data archiving (e.g. WDC-MARE,Sismer)*

Pangea

- *Will the DM provide access to observatory data via Sensor Observation Service ?*

No, the observatory is not online.

- *Will the sensor specification be available in SensorML ?*

I am not familiar with this.

- *Do you plan to support other parts of the Sensor Web Enablement protocols ?*

I am not familiar with this, no idea what it means.

LIDO :

- Where will the data delivered by the DM be stored initially ?

Initially, data will be stored at the institutions responsable for the acquisition.

- Is the data accessible via the internet, if so where ?

Data (and resulting information) will be accessible through the internet, http://www.listentothedeep.org

- Which data center will be responsible for the long-term data archiving (e.g. WDC-MARE,Sismer)

WDC-MARE, at least for data analysis results. Due to its size, raw acoustic data might be stored differently.

- Will the DM provide access to observatory data via Sensor Observation Service ?

This is the intention.

- Will the sensor specification be available in SensorML ?

It should be.

- Do you plan to support other parts of the Sensor Web Enablement protocols ?

Yes we do, especially the alert services once the standard has been clearly defined, otherwise the fall-back alternative is RSS.

Appendix II: WP9 Meeting Vilanova i la Geltrú

Following is a summary of the results of a WP9 meeting (in the frame of the exchange of personal within the ESONET) concerning data management practices.

Workpackage 9 - Data Management Meeting Polytechnic University of Catalonia, Vilanova i la Geltrú, February 24-25, 2010.

Present:

Andree Behnken (Marum)

Christian Curtil (CPPM-CNRS)

Aurelien Maurin (CPPM-CNRS)

Cécile Robin (Ifremer) (remotely)

Mike van der Schaar (LAB-UPC)

Data Archiving

For ANTARES it is not foreseen to start archiving to the WDC-Mare before November, when the new line under the Albatros project might become operational.

For LIDO it has been decided to only store analysis results (this includes noise measurements and other statistical information, occurrences of cetaceans, shipping information, etc). Raw data will be considered once it becomes clearer what kind of data might be interesting for long term storage.

The granularity of the data moved to the WDC-Mare will be daily. If there is a change in meta-data information during the day, then the data will be split into one set per meta-data file. The granularity for analysis is much smaller; in the case of LIDO the data set will contain a list of analysed segments of 17 seconds.

While it would be preferred to have human quality control on the data before it is transferred to the WDC-Mare, this might not be feasible. In the case of LIDO, there will be automated detection of saturation, which will be documented in the segment analysis in a separate field (this does not automatically mean that the whole segment was unusable). Another indicator for poor quality are the noise measurements taken on each segment. While cetacean detection might be unreliable in high noise conditions, the noise information itself can still be considered valuable (and thus suitable for archiving).

Data Formats

Datasets that are moved to the WDC-Mare will either include meta-data and data in the same file, or the two will be divided in two separate files; meta-data will describe:

- sensor information (likely through a link to a specific sensor ml file)
- procedure information
- algorithms used for postprocessing results

- data descriptions (e.g. data types)
- institute/citation information
- time information

Regarding time information, each processed segment from LIDO will carry its own timestamp; the meta-data timestamp will have the period that is covered by the information in the accompanying data set - without taking into account possible missing segments (gaps).

The intention is to describe everything in OGC (XML) standards and not to include something like Netcdf that would require another library to be linked with the code.

As using the O&M procedure element for describing meta-data information (describing the processing done to obtain specific outputs) will upset the SensorML people, we will look into using the ProcessModel for non-physical or pure processes from SensorML. It is important for ease of use that there is a text-field where we can simply put a rough description of the meaning of a value and how it was obtained (although a latex-field might be nice too for some parts of the algorithms), without having to specify much more.

To describe the data types, if i understand Cécile correctly (email copied verbatim below), we can use the namespace from http://www.w3.org/TR/xmlschema-2/ in the meta-data description to describe the data types in detail (i believe that e.g. something like xs:integer is preferred over swe:Count, as a 'count' has a specific incremental context). Use of the xs namespace seems to make XML|Text|Binary blocks unnecessary, at least until raw data is to be encoded.

It is noted that it might be difficult to ensure correctness of the (meta-)data encodings, without proper verification. For the hardware descriptions, ANTARES will rely on the output of ESONET WP2 to provide hardware templates that can be filled in.

Data Transfer

At this moment, a real-time connection between ANTARES/LIDO and the WDC-Mare is not considered. In the future the Pangaea portal may provide access to real-time datastreams served from the data providers.

For LIDO data, we will try to set-up an SOS server based on the Oostethys-perl-server, which seems to be fairly easy to adapt to the LIDO-case. The only requests that will be supported are likely to be GetCapabilities, DescribeSensor and GetObservation. Initially, both requests and answers are expected to follow a strict template; requests that would not strictly follow the template would be rejected with an error (this means that only requests for complete data segments will be accepted; thinking of this, i am not sure if the daily-data-set file for the WDC-Mare would be returned as a whole inside a GetObservation request, to be looked at…). The error message might be provided using owsExceptionReport.xsd, this needs to be looked at as well (inclusion of yet another namespace).

For ANTARES data, a first approach that can be supported relatively easy is to upload the data to an (S)FTP server at the WDC-Mare, but the interest will be to await the results of the LIDO - WDC-Mare SOS test and to follow that approach.

Message-ID: <4B8C0347.7050109@ifremer.fr>

Date: Mon, 01 Mar 2010 19:11:19 +0100
From: Cecile ROBIN <Cecile.Robin@ifremer.fr>
To: Mike Vanderschaar/UPC <Mike.Vanderschaar@upc.edu>
CC: Christian Curtil <curtil@cppm.in2p3.fr>,
Michel Andre/UPC <Michel.Andre@upc.edu>,
Andree Behnken <abehnken@marum.de>,
Jean-jacques Destelle <destelle@cppm.in2p3.fr>,
Eric Delory <edelory@ieee.org>, eric.delory@dbscale.com,
maurin@cppm.in2p3.fr, Robert Huber <rhuber@uni-bremen.de>
Subject: UPC WP9 Workshop - reply to few points

Dear all,

Thank you for the meeting again.
Reply top few questions:

1) Data Type
1.1) Basic numerical types definition
this types are actually not defined in the gml namespace but in W3C (usually using the 'xs' prefix) see http://www.w3.org/TR/xmlschema-2/

1.2) Swe types
Possible numerical values in swe are (see AnyNumerical in SimpleTypes.xsd): "swe:Count" "swe:Quantity" "swe:Time"

"swe:Quantity" is associated with a 'xs:double' value
"swe:Count" is associated with a 'xs:integer' value
"swe:Time" is associated with a 'xs:double' value

1.3) xs:double definition:
see
http://www.w3.org/TR/xmlschema-2/#built-in-primitive-datatyp
es For example, -1E4, 1267.43233E12, 12.78e-2, 12 , -0, 0 and INF are all legal literals for double.

2) O&M data encoding for the data block
O&M data encoding is specified by the "swe:encoding" element. Choice is - TextBlock: (specify tokenSeparator, decimalSeparator, blockSeparator) - BinaryBlock: ... - StandardFormat: (specify mimeType) - XMLBlock

I think XMLBlock should not be allowed (too verbose), I am not sure how to use the standardFormat, however:

a TextBlock can contain any -1E4, 1267.43233E12, 12.78e-2, 12 type of values but NO Hexadecimal values

a BinaryBlock can contain
- base64 elements, hopefully this should correspond to the IEEE 754 floating-point "double format" bit layout - raw - hex

Do we want to support binary block?

3) An another important point is the CHARSET of the file The real Charset of a file can't be detected easily: An xml file can specify 'utf8' in its header but be encoded in windows cp-1250 or in western ISO... This problem can arise when saving a file in a non-xml context (using a simple text editor,...) Note that there is not such problem with netcdf...

4) O&M featureOfInterest - 52degreeN
I confirm that encoding the id of the featureOfInterest within the dataBlock is a Choice from 52degreeNorth. See for example
http://schemas.opengis.net/om/1.0.0/examples/weatherObservation.xml or SOS spec p.51 (InsertObservation example) where latitude and longitude are given inline within the data and not encoded inside the featureOfInterest

5) O&M Process
I am still very much in favour to use only sensorML types of processes The reason is om:Process can't be a standalone document, it does not correspond to any SOS operation whereas the sensorML process does

We'll see if there are other opinions on that subject

cheers

--
Cécile

Appendix III: XML/schema examples from LIDO

A. Schema file for the analysis results of LIDO, including documentation on the analysis process

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="http://www.w3.org/2001/XMLSchema"
  xmlns:gml="http://www.opengis.net/gml" xmlns:gmd="http://www.isotc211.org/2005/gmd"
  xmlns:om="http://www.opengis.net/om/1.0" targetNamespace="http://www.listentothedeep.net/"
  elementFormDefault="qualified" attributeFormDefault="unqualified" version="1.0.0"
  xmlns:lido="http://www.listentothedeep.net/">

<annotation><documentation>
  A full explanation of the LIDO framework can be found in \cite{XXX}.
</documentation></annotation>

  <xs:group name="probability">
    <annotation><documentation>This is a list of probabilities for the occurence of a specific acoustic event or presence of a marine mammal in the analysed segment. </documentation></annotation>
    <xs:all>
      <xs:element name="dolphin">
        <annotation><documentation>
          Dolphin presence is measured based on the occurence of whistles in the segment. The whistle detector has been decribed in \cite{XXX}. A single strong whistle will already give a high probability as there are few other sources that produce these signals.
        </documentation></annotation>
        <simpleType><restriction  base="integer"><minInclusive value="0"/><maxExclusive value="100"/></restriction></simpleType>
      </xs:element>
      <xs:element name="shipping">
        <annotation><documentation>
          Shipping presence is measured based on a combination of the presence of a high number of broadband impulse sounds, constant tonal sounds for the duration of a segment and strong background noise.
        </documentation></annotation>
        <simpleType><restriction  base="integer"><minInclusive value="0"/><maxExclusive value="100"/></restriction></simpleType>
      </xs:element>
      <xs:element name="spermwhale">
        <annotation><documentation>
          The presence of sperm whales is measured based on impulsive sounds with peak energy around 15 kHz. A neural network classifier (\cite{XXX}) is used to distinguish between sperm whale and shipping impulses.
        </documentation></annotation>
        <simpleType><restriction  base="integer"><minInclusive value="0"/><maxExclusive value="100"/></restriction></simpleType>
      </xs:element>
    </xs:all>
  </xs:group>

  <xs:element name="segment">
    <annotation><documentation>A segment is the time interval that is analysed at one time and described in the result file.</documentation></annotation>
    <xs:complexType>
      <xs:all>
        <xs:element name="platformName" type="string">
          <annotation><documentation>The name of the platform where the data comes from. The sensor description should already include information on the specific set-up at this location.</documentation></annotation>
        </xs:element>
        <xs:element name="sensorDescription" type="anyURI">
          <annotation><documentation>
            This is a link to the description (a SensorML file) of the hardware that is used at the site.
          </documentation></annotation>
        </xs:element>
        <xs:element name="timestamp" type="dateTime">
```

```
          <annotation><documentation>
            This is the timestamp at which the data was recorded, originally it was a UNIX timestamp which is converted to the XML data type.
          </documentation></annotation>
        </xs:element>
        <xs:element name="RMS">
          <simpleType>
            <annotation><documentation>
              This value measures the noise in dB RMS re 1 uPa for the whole segment. It can be safely assumed not to exceed 300 dB.
            </documentation></annotation>
            <restriction base="double"><minInclusive value="-100"/><maxInclusive value="300"/></restriction>
          </simpleType>
        </xs:element>
        <xs:element name="peakLevel">
          <simpleType>
            <annotation><documentation>
              This is the peak level measured in dB re 1 uPa found in the segment.
            </documentation></annotation>
            <restriction base="double"><minInclusive value="-100"/><maxInclusive value="300"/></restriction>
          </simpleType>
        </xs:element>
        <xs:element name="probabilities">
          <xs:complexType>
            <xs:group ref="lido:probability"/>
          </xs:complexType>
        </xs:element>
      </xs:all>
    </xs:complexType>
  </xs:element>
</xs:schema>
```


B. Encoding of a single segment of LIDO data


```
<?xml version="1.0" encoding="UTF-8"?>
&lt;segment xmlns="http://www.listentothedeep.net/" xmlns:sml="http://www.opengis.net/sensorML/1.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.listentothedeep.net/ file:/F:/XML/lido.xsd">

  <sensorDescription>http://some.location.net</sensorDescription>

  <platformName>ESONET PLATFORM</platformName>
  <RMS>115.67</RMS>
  <peakLevel>125</peakLevel>
  <timestamp>2010-03-01T12:00:00+01:00</timestamp>
  <probabilities>
    <dolphin>0</dolphin>
    <shipping>20</shipping>
    <spermwhale>1</spermwhale>
  </probabilities>
&lt;/segment&gt;
```


C. Example of GetObservation answer, currently based on Oostethys example, to be defined with WDC-MARE


```
<om:ObservationCollection xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.opengis.net/om/1.0 http://schemas.opengis.net/om/1.0.0/om.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:om="http://www.opengis.net/om/1.0"
  xmlns:gml="http://www.opengis.net/gml" xmlns:swe="http://www.opengis.net/swe/1.0.1"
  xmlns:lido="http://www.listentothedeep.net/"
  gml:id="OFFERING_DATA">

  <gml:description/>
```

```xml
  <gml:name/>
  <gml:boundedBy>
    <gml:Envelope srsName="urn:ogc:def:crs:EPSG:6.5:4326">
      <!-- overwrite these with your actual offering ROI -->
      <gml:lowerCorner/>
      <gml:upperCorner/>
    </gml:Envelope>
  </gml:boundedBy>
  <om:member>
    <om:Observation>
      <gml:description/>
      <gml:name/>
      <gml:boundedBy>
        <gml:Envelope srsName="urn:ogc:def:crs:EPSG:6.5:4326">
          <!-- overwrite these with your actual offering  ROI -->
          <gml:lowerCorner/>
          <gml:upperCorner/>
        </gml:Envelope>
      </gml:boundedBy>
      <!-- Observation time -->
      <om:samplingTime/>
      <!-- Sensor description (SensorML) -->
      <om:procedure xlink:href="urn:to:you:unique:sensorOrProcedure"/>
      <!-- Observed Property, can be multiple -->
      <om:observedProperty>
        <swe:CompositePhenomenon dimension="1" gml:id="BUOY_OBSERVABLES">
          <gml:name>"wer"</gml:name>
          <swe:base/>
          <swe:component/>
        </swe:CompositePhenomenon>
      </om:observedProperty>
      <om:featureOfInterest xlink:href="urn:ogc:def:object:feature:FOI_URI"/>
      <!-- Result Structure, Encoding, and Data -->
      <om:result>
        &lt;lido:segment xmlns="http://www.listentothedeep.net/" xsi:schemaLocation="http://www.listentothedeep.net/ file:/F:/XML/lido.xsd">
          <probabilities>
            <dolphin>1</dolphin>
            <shipping>1</shipping>
            <spermwhale>1</spermwhale>
          </probabilities>
          <sensorDescription>http://nowhere.com</sensorDescription>
          <platformName>ESONET PLATFORM</platformName>
          <RMS>115.67</RMS>
          <peakLevel>125</peakLevel>
          <timestamp>2010-03-01T12:00:00+01:00</timestamp>
        &lt;/lido:segment&gt;
      </om:result>
    </om:Observation>
  </om:member>
</om:ObservationCollection>
```