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

D43-D44 - Data Infrastructure Productive Version

ESONET Knowledge Base

Due date of deliverable: month 42
Actual submission date: month 43

Start date of project: March 2007
Duration: 48 months

Organisation name of lead contractor for this deliverable: Participant # 6, KDM, Germany
Lead authors for this deliverable: Robert Huber, Andree Behnken, Thierry Carval, Eric Delory & Cecile Robin

Revision [4, 3.11.2010]

---

<table>
<thead>
<tr>
<th>Dissemination Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PU Public</td>
<td>X</td>
</tr>
<tr>
<td>PP Restricted to other programme participants (including the Commission Services)</td>
<td></td>
</tr>
<tr>
<td>RE Restricted to a group specified by the consortium (including the Commission Services)</td>
<td></td>
</tr>
<tr>
<td>CO Confidential, only for members of the consortium (including the Commission Services)</td>
<td></td>
</tr>
</tbody>
</table>
Introduction

The productive ESONET data infrastructure is the result of intensive networking and cooperative efforts including both observatory and sensor experts contributing knowledge on observatory architectures, as well as IT and data management specialists experienced in scientific knowledge and data handling. This fruitful cooperation and ‘bottom-up’ approach led to the current standard-based infrastructure architecture as well as its productive implementation.

The data infrastructure for ESONET is designed as a distributed system. Both observatory data as well as data archiving services are already provided by several observatory nodes and data centres. Therefore, the main challenge for the ESONET data infrastructure was to provide a technical architecture based on international standards to implement data management policies and work flows. ESONET also has developed an online knowledge base for general information about the observatory system and how ESONET data management fits into a larger context. Beneath common standards for metadata description and exchange such as OAI-PMH and ISO19139, ESONET has chosen to implement core standards of the Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) suite of standards, namely the OGC standards SensorML, Sensor Registry, Catalogue Service for Web (CS-W), Sensor Observation Service (SOS) and Observations and Measurements (O&M).

Fig. 1: Overview of the main standard components used for the ESONET technical infrastructure.

However, in order to be useful for as many applications as possible, each of the above mentioned SWE standards rather represents a generic and abstract framework instead of detailed implementation rules. They intentionally allow much interpretative freedom and have several implementation approaches. Therefore, to ensure internal compatibility, so called application profiles have to be defined and accepted by OGC.
The definition of these common profiles was one of the major challenges for ESONET and it soon became clear that this had to be done in cooperation with other major European initiatives. This led to an intensive and successful collaboration with e.g. EuroSites, OceanSites and SeaDataNet which resulted in several common specifications. All these standardisation efforts have of course been carried out together with the team of workpackage 2.

The formation and collaboration of this cross-project, cross-workpackage data management group surely is one of the most remarkable results of ESONET. This informal group truly consists of enthusiastic specialists and their joint efforts as drafted above were only possible in the friendly atmosphere among those people involved.

The most obvious application of the standards mentioned above is the ESONET knowledge base, the general information platform for ESONET. It serves data from ESONET sites and observatories via the internet. For example, it uses OAI-PMH to harvest metadata from contributing archives and uses SOS to retrieve the latest data from observatories. It uses O&M to visualise data and SensorML to display observatory information.

The technical platform managing the metadata catalogue of contributing data archives is panFMP (PANGAEA Framework for Metadata Portals, Schindler and Diepenbroek 2008). Furthermore, it provides access to sensor metadata via the ESONET sensor registry and offers access to real-time data delivered by the mentioned SOS servers, as well as to archived data. Additionally, it allows to access data from third party data centres which support the standards described above and are of relevance for ESONET.

The following document describes the ESONET implementation of the above mentioned OGC standards as integral part of the ESONET data infrastructure as well as the knowledge base representing the user interface to the ESONET data infrastructure.

The ESONET Data Infrastructure

Distributed data archives and metadata exchange

Data archiving and metadata standards:

Trustworthy data centres are the backbone of the ESONET data archiving strategy. According to the ESONET data policy, World Data Centres and other certified data centres like National Data Centres (IOC/IODE) ideally are the responsible bodies for the long term archiving of the project’s data. This ensures long term availability of ESONET data and allows the definition of high level architectures for distributed data management.

The ESONET distributed data archive infrastructure requires a highly standardised approach to data archiving as well as meta description of available data sets. This includes efforts towards a common granularity of data (the definition of a data set) as well as the usage of common metadata formats. Harmonisation and standardisation of data and metadata is therefore essential. The choice of data and metadata exchange standards in the field of marine research was already described in detail in the ESONET data and information plan (D9 & D70). However, besides a common technical architecture, a trustworthy distributed data system also needs a social contract between data providers and centres. Therefore, the rules and responsibilities of both data providers as well as data archives additionally have been defined in detail in deliverable D9.

Compatibility of metadata is most important. Currently, the ISO 19115¹ and its XML implementation defined in ISO 19139 is the most important metadata standard for

¹ ISO 19115 Geographic information - Metadata http://www.iso.ch
geographically referenced metadata description. To support data centres which have not yet adopted the new ISO 19139, other XML based standards are also accepted. For example, the DIF\(^2\) format is one of these standards. It was originally proposed by the GCMD and is widely accepted within the marine data community. The DIF format is used for example within the IODE initiated MEDI\(^3\) portal.

**Data integration and exchange protocols:**

Beneath the harmonisation and standardisation of metadata, the choice of metadata exchange standards is of special importance and was already described in detail in the ESONET data and information plan (D9). For ESONET, the Open Archives Protocol\(^4\) (OAI-PMH) was chosen. OAI-PMH, a popular exchange protocol for metadata on physical as well as digital objects, is well suited. It is simple and easy to implement but very flexible with regard to metadata formats. For example, most libraries offer this protocol to exchange their library catalogues metadata. OAI-PMH was recommended by the IODE\(^5\) for marine science related metadata exchange\(^6\) and is used in several scientific data portals (scientific data: SEDIS: [http://sedis.iiodp.org](http://sedis.iiodp.org); EUR-OCEANS: [http://www.eur-oceans.org](http://www.eur-oceans.org); museums: BAM: [http://www.bam-portal.de](http://www.bam-portal.de), Europeana: [http://www.europeana.eu](http://www.europeana.eu)). OGC CS-W\(^7\) is another example for a standardised network protocol for distributed searches on metadata in the geo-scientific world. While OAI-PMH is best suited for metadata harvesting and is used to integrate external data archives, OGC standards will be used to link ESONET data to international initiatives such as GEOSS. The Open Geospatial Consortium (OGC) standards implementation will provide open services for ESONET discovery, visualization and data download.

Several partners have enabled their repositories for either ISO or DIF format. Among others, the Ifremer offers metadata in DIF format, the WDC-MARE in DIF as well as ISO format and the INGV has provided DIF formatted metadata. Both the WDC-MARE as well as the Ifremer offer an OAI-PMH interface and the INGV is heading towards a similar solution.

Some partners already contribute metadata to SeaDataNet, e.g. the HCMR. We therefore initiated a vivid cooperation with the SeaDataNet project which has led to major improvements within the ESONET metadata exchange workflow. For example, data from the POSEIDON observatory already has been submitted by HCMR to SeaDataNet. Fortunately, SeaDataNet now provides standardised metadata exchange protocols which are fully compatible to the ESONET data infrastructure such as OAI-PMH or CSW. Furthermore, SeaDataNet provides ESONET compatible metadata in ISO19139 format. After some initial and very successful interoperability tests in July 2010 we can make use of the SeaDataNet infrastructure and harvest POSEIDON metadata for inclusion into the ESONET portal.

**Sensor and observatory information management**

**Catalogue Service for the Web (CS-W)**

The OGC Standard CS-W provides the methodology to publish and access digital catalogues of metadata for geospatial data, services, and related resource information. It defines for example a suite of standardised requests in XML format to retrieve geospatial metadata, e.g.

---


\(^3\) Marine environmental data inventory http://ioc.unesco.org/medi/

\(^4\) http://www.openarchives.org

\(^5\) IOC Committee on International Oceanographic Data and Information Exchange

\(^6\) Reports of Governing and Major Subsidiary Bodies, IOC Committee on International Oceanographic Data and Information Exchange, Eighteenth Session Kursaal, Ostend, Belgium, 26-30 April 2005, Action sheet item 6

\(^7\) Open Geospatial Consortium Catalogue Service for Web; http://www.opengeospatial.org/standards/cat
via the common internet http protocol. The development of the Sensor Registry and the
catalogue service have been closely related. Within the ESONET data infrastructure, a CS-W
service allows access to SensorML data which is stored in the Sensor Registry. Where the
Sensor Registry is the repository where all sensor related information is stored, the catalogue
service CS-W provides the possibility to access this information in a standardized way.

The CS-W protocol is very flexible and allows the support of various metadata formats.
However, to ensure standard compatibility, so called application profiles have to be defined
and accepted by OGC. Currently, CS-W services are commonly used to provide access to
metadata in ISO 19115 format only, such as map data or geocoded scientific data.
Unfortunately, there is no CS-W implementation available which supports SensorML.

Therefore, we have implemented a first demo version of a CS-W server in PHP to support
SensorML. Many other OGC protocols share some basic commands and functionalities,
which are described in the OGC Web Services Common Standard (OGC 06-121r9). These
common services can be used to write reusable software code. Because UPC was starting to
implement a OGC SOS server also in PHP, we have joined efforts and initiated a common
code base for those common services which we then used for specific SOS and CS-W
implementations. This common code base was published as Open Source under an Apache
License at Google Code and can be downloaded at: http://code.google.com/p/esonet-ogc.

As a pragmatic approach the ESONET CS-W server is based on a mapping of SensorML
elements to the ‘core queryable elements’ defined in the CS-W standard documentation.

<table>
<thead>
<tr>
<th>Common queryable element</th>
<th>SensorML Xpath</th>
</tr>
</thead>
</table>
| Title                    | As the name element is not obligatory and there are several possibilities to express a sensor’s name, Xquery will check the following possibilities:  
  if a longname exists:  
  otherwise check if a shortname exists:  
  otherwise check if the gml:name exists:  
  /sml:SensorML/sml:member/sml:System/gml:name  
  last possibility:  
| Abstract                 | /sml:SensorML/sml:member/sml:System/gml:description/AnyText XML Fulltext |
| Modified                 | XML modification date |
| Type                     | dataset |
| BoundingBox             |  
  ‘upper lat’:  
  ‘upper lon’:  
  ‘lower lat’:  
This coordinated approach allowed to implement a CS-W on top of the Sensor Registry in a comparably short period of time. A first public beta version of the productive CS-W which supports querying on these common elements can be visited at: http://dataportals.pangaea.de/esonet/ogc. A simple client that allows entering XML queries is available at http://dataportals.pangaea.de/ogc/client.php.

The following XML can be used as a sample request which should return the capabilities of the CS-W server:

```xml
<csw:GetCapabilities service="CSW" xmlns:csw="http://www.opengis.net/cat/csw" xmlns:ows="http://www.opengis.net/ows">
  <ows:AcceptVersions>
    <ows:Version>2.0.0</ows:Version>
  </ows:AcceptVersions>
  <ows:AcceptFormats>
    <ows:OutputFormat>text/xml</ows:OutputFormat>
  </ows:AcceptFormats>
</csw:GetCapabilities>
```

**Sensor Registry**

The Sensor Registry Registration Interface (SRI)’s objective was to plan, develop and deliver a process which facilitates the registration and register maintenance of a network of instruments and sensors with the goal that the collected observations be packed with relevant sensor metadata encoded in an open standard format. SensorML was chosen on the basis of ESONET internal consultation and international acceptance. The overall goal was to minimize human intervention by facilitating the discovery of sensors across the global network.

The SRI web interface permits the input of metadata from a deployed sensor or multi-sensor instrument (CTD, ADCP, etc.). The sensors are converted into smart sensors as they provide sufficient information for machines to process observations and deliver information. A detailed description of the SRI functionalities is provided in Deliverable D72.
Fig 2: Some SRI Web service screenshots, detailed information on the Sensor Registry can be found in deliverable D72

Sensor Model Language (SensorML)

The Sensor Model Language (SensorML) is an OGC standard which defines a XML format for describing sensor systems and processes. A SensorML document provides information needed for discovery of sensors, location of sensor observations, processing of low-level sensor observations, and listing of sensor system operations that can be invoked by a client process. SensorML can contain a variety of information describing a sensor or system such as the capabilities and characteristics of a sensor, e.g. its physical measures (length, weight etc.). The format can be used to describe what a sensor actually measures and what it returns as output parameters. Further, meta-information such as the name of a sensor, its manufacturer or responsible parties can be described by SensorML as well as the sensor documentation including calibration data. Furthermore, SensorML provides the possibility to describe sensor communication protocols and other interfaces.

Within ESONET, SensorML is used for three purposes. First, it is needed for the SOS protocol as a formal description of the sensor characteristics. Second, SensorML is fed to the SensorRegistry where it is stored and can be queried using the OGC catalogue service CS-W. It is also used within the ESONET knowledge base to give the public an overview on the
sensors using XSLT transformations to transform the XML format to a human readable web format.

SensorML is a very generic standard. It can be considered as a framework to define community profiles. In cooperation with OSIRIS, OceanSites and EuroSites we have therefore defined a common format for European marine observatories. This standardisation effort included the definition of granularities for SensorML entities which ESONET applied for the creation of XML files for submission to the Sensor Registry. Three main sensor or observatory entities have been identified which represent different levels within the ESONET sensor network framework:

- **Observatory / Site level**
- **Deployment level**
- **Instrument level**

These levels are a unique feature for marine observatories and the main reason why major efforts for the specification of an ESONET SensorML profile became necessary. The data management group has agreed that the instrument level will be the main ESONET sensor entity used within sensor registry. However, the deployment level will additionally be stored within the sensor registry.

A detailed documentation of the ESONET SensorML profile has already been submitted in deliverable D42. It has been defined in close cooperation with WP2 which intends to use the profile as a basis to define a specific application profile for submission to OGC. The ESONET SensorML specification has additionally been defined in a XML Schematron file which is available at the ESONET homepage.

### Core standards for observatory data transfer

**Observation and Measurements (O&M)**

ESONET uses the **SOS** service to harvest data for automated data flow from observatory nodes to a selected data archive. OGC standards are used to generate International Standardization Organization (ISO) 19139 compliant metadata descriptions for the harvested data. O&M and ISO data is then uploaded to an appropriate data centre and ingested. Currently, the World Data Center for Marine Environmental Sciences (WDC-MARE) as well as Systèmes d'Informations Scientifiques pour la Mer (SISMER) are implementing routines to acquire data from ESONET observatories and are capable of long-term data archiving services. In addition, both data archives provide data curatorial procedures including quality control for manual or semi-automatic upload of data to the archive. Each ESONET data archive further provides XML metadata files on each data set in either ISO 19139 or Global Change Master Directory - Directory Interchange Format (GCMD DIF) format and offers metadata exchange via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface.

**Sensor Observation Service (SOS)**

The OGC standard **Sensor Observation Service (SOS)** allows direct access to the data of a cabled ESONET observatory. SOS is a standard web service interface for requesting, filtering, and retrieving observations and sensor system information. It allows to e.g. directly retrieve the latest data collected by a sensor or to use the service to harvest or download data for example on a daily or weekly basis. Observatory data provided by the SOS is ideally
represented in the OGC standard format **Observations & Measurements** (O&M) (see above). Beneath providing a data access protocol, a SOS enables a sensor or observatory to describe itself by providing a SensorML document containing the sensor’s metadata. Therefore, the observation service will closely relate to the Sensor Registry as well as the catalogue service described above.

Compatibility with other international observatory systems was a priority during the development of the SOS. Some other initiatives also use this standard; however, due to the already mentioned generic definition of OGC standards, each community may already use their own application profiles which may only be partially compatible. From the beginning we have therefore cooperated with other initiatives, for example with members of the fp6 OSIRIS\(^8\) and the SANY project. The OSIRIS team used the 52°North\(^9\) SOS implementation. 52°North currently adopts this package for the Tsunami Early Warning System. Most important was the cooperation with the EuroSites and OceanSites project led by Ifremer. This excellent cooperation resulted in a common implementation of OGC standards which can be regarded as the basis for a European marine observatory standardisation effort which will be continued within EMSO.

Two new SOS software implementations, based on these standardisation efforts, have been developed and are used at ESONET observatories. Further, several third party SOS software packages are now on the market, which potentially can be used by marine observatories. In addition to the new developments in cooperation with EuroSites we therefore also have analysed alternative systems:

As part of ESONET’s data management work package, three different SOS implementations have been tested and evaluated with regard to importing sensor data into PANGAEA.

**52°North SOS**
The 52° North SOS is a Java package and the only one that supports transactional operations (Table 2). It requires the use of a specific database scheme which is shipped as part of the package. This makes it somewhat inflexible, since most sensor operators will have their own database system for data storage. Of the tested implementations, this one is the most complex.

**PySOS / Perl SOS**
These were both implemented by OOSTethys ([http://www.oostethys.org](http://www.oostethys.org)) and consist of simple Python / Perl scripts. They have been developed for specific OOSTethys purposes and can not be used "out-of-the-box". In order to use these servers for PANGAEA purposes, recoding and adaptation would be necessary.

---


\(^9\) [http://52north.org](http://52north.org)
The following table gives an overview on the tested implementations.

<table>
<thead>
<tr>
<th>Organisation</th>
<th>52°North</th>
<th>OOSTethys</th>
<th>OOSTethys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>52°North SOS</td>
<td>Python SOS server (PySOS)</td>
<td>Perl SOS server</td>
</tr>
<tr>
<td>Program language</td>
<td>Java</td>
<td>Python</td>
<td>Perl</td>
</tr>
<tr>
<td>Mandatory operations:</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>GetCapabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GetObservation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DescribeSensor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transactional operations:</td>
<td>x</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RegisterSensor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InsertObservation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional operations:</td>
<td>x</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GetFeatureOfInterest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GetResult</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: SOS implementations.

Since the tested implementations are not completely satisfactorily, it is planned to develop a JAVA based generic solution that, accompanied by a SOS client, can be (re)used for data retrieval from various providers and data import into PANGAEA without the necessity of recoding the software.

The Ifremer SOS server

Ifremer has developed, set up and tested an OGC SOS server in 2009 and 2010 focusing in a first step on time-series observations. On this server, the description of the site of observation and its meta-data is performed with an OGC-SensorML file. The data access is performed through OGC-SOS standard protocol. Since October 2010, time-series observations are available on the internet from the Ifremer SOS server. A graphic client of the SOS server was developed to allow interactive data display and download (see figure 1). The observatory meta-data (site, platform, deployment and sensor) are displayed on a web-html page, or distributed as xml sensorML files (see figure 4).
Vertical profiles observations are now under tests on Ifremer server and should be available by the end of 2010. We are planning to gradually extend it to other types of observations: images, video and acoustics.

The software development was carried out in cooperation with the Geomatys Company which implemented the SOS-SWE server at Ifremer. The name of the server software is Constellation and is based on Java-J2EE projects such as GeoServer, GeoTools and GeoAPI. More information can be found here: [http://www.geomatys.fr/geomatys/main.jsp?adress=Geotools.jsp](http://www.geomatys.fr/geomatys/main.jsp?adress=Geotools.jsp).

Until now, several EuroSITES sites are available from the SOS server among those CIS, ESTOC, PAP, E1M3A and PYLOS. These services are currently tested and will successively be included to the ESONET data portal.
The ESONET Open Source SOS server

During our tests with available SOS servers it turned out that some ESONET partners would need a light weight SOS solution which should be easier to adopt to an observatory’s need as well as to available technical platforms than the available solutions.

Initiated by the UPC and Uni-HB we therefore started to implement a simple PHP based SOS server, which was decided to be offered as Open Source. The code is based on the ESONET common OGC PHP classes we also used for the CS-W server classes mentioned above and is publicly available for download at the ESONET Google Code repository at http://code.google.com/p/esonet-ogc/ under an Apache 2.0 licence.

The first implementation of the ESONET Open Source SOS now delivers hydrophone data from the ANTARES observatory obtained during the LIDO demo mission and is available at: http://147.83.13.179:180/sos/sos.php (the URL may be subject to change after the testing phase). The SOS server currently is in a beta stage and intensively tested.

Data access via this SOS server is also intended to be used for semi-automatized data ingestion and archiving at the WDC-MARE. It is therefore a major component of the future workflow for archiving data from the LIDO demo mission at the WDC-MARE.

WDC-MARE SOS data harvesting client

Figure 5: Data harvesting schema describing the flow of data from the observatories’ SOS to the WDC-MARE data archive system PANGAEA
The WDC-MARE SOS client is designed to send periodic GetObservation requests to SOS servers, in order to obtain data provided by those servers, store the server’s xml responses, write the data to text files and finally create an import file conforming to the ascii import format used for the WDC-MARE’s archive system PANGAEA.

Initially, these files will be additionally double checked by a team of experienced data curators before they enter the WDC-MARE’s data upload batch. However, in the future these import files can be used to import and store data in the PANGAEA geoscientific database without further processing. Thus, the system will allow the fully automatized archiving of marine observation data based on standards defined within the OGC SWE family of standards.

The harvesting client is a command line tool, which uses an xml configuration file holding all necessary operational information. This file is used to configure the client on startup and defines, amongst others, the URL of the SOS server and specifics for PANGAEA import file generation. The harvesting procedure is performed once per day and extracts the relevant data from the SOS server’s xml responses and appends this new data to a temporary data cache which is a simple comma separated text file. Once the client recognizes the end of each harvesting cycle, a new harvesting process is initiated and the current data cache file is closed and transferred to import preparation process. The import cache file is renamed to represent the current month of the year (yyyy_mm), its contents are used to create the import file for this month, and a new temporary cache file is created and used to write all subsequent data to. Data from the client configuration XML file are used to generate the metadata (header) section of the import format; the data section is attached below. The result is a WDC-MARE proprietary import file format such as:

```
/*DATA DESCRIPTION:
   Author: 36736, 36735
   Title: LIDO Hydrophone Data, Mediterranean Sea (2010_09)
   Reference: 23622
   Export Filename: lido_2010_09.txt
   Event: ANTARES_1
   PI: 36736
   PARAMETER: 1599 * PI: 36736 * FORMAT: yyyy-MM-dd'T'HH:mm:ss
   91468 * PI: 36736 * METHOD: 6916 * FORMAT: ##0.000
   91469 * PI: 36736 * METHOD: 6916 * FORMAT: ##0.000
   91470 * PI: 36736 * METHOD: 6916 * FORMAT: ##0.000
   Project: 4133
   Topologic Type: 2
   Source: 3038
   Status: validated
*/

Event label 1599 91468 91469 91470 91471

ANTARES_1  2010-09-01T02:00:00:03  0  90.63999389648 112.88999938965 0
2010-09-01T02:00:20  0  96.59999847412 139.77000427246 0
2010-09-01T02:00:36  24 107.75 141 0
2010-09-01T02:01:53  0  90.26000213623 113.38999938965 0
2010-09-01T02:01:10  0  90.459999984743 115.87000274658 0
2010-09-01T02:01:26  0  90.339996337891 115.55999755859 0
2010-09-01T02:01:43  0  90.089996337891 118.04000091553 0
2010-09-01T02:02:00  0  90.110000610352 116.16000366211 0
2010-09-01T02:02:17  2  99.019996643068 140.02000427246 0
2010-09-01T02:02:34  26 107.41999816895 141 0
```

Both sections represent the import format for PANGAEA. The import files are stored in a dedicated data curator directory and are ready to be used for data import into Pangaea at any desired time.

If a new import file was generated, a WDC-MARE data curator will be notified of this event via an automated email message generated by the client. Likewise, the client will also generate a message in case of any errors.
The WDC-MARE data harvesting client is currently tested and will be used to automatically transfer e.g. LIDO hydrophone data directly from the observatory to the data archive. We expect to be able to present the first archived data from LIDO's NEMO platform at the ESONET data portal soon after this observatory will be deployed in November 2010.

**ESONET SOS generic client**

To include SOS delivered data to the ESONET data portal, the WDC-MARE has implemented a generic SOS web client. The challenge here was to enable the client to understand different SOS response formats as it turned out that some partners already had different SOS server solutions in use or were about to test such implementations which are not 100% compatible to the ESONET O&M profile.

The ESONET generic SOS client is able to interpret responses from 52°North SOS server, the OOSTheys Python SOS server, the Ifremer SOS server as well as the ESONET open source SOS server. Whereas the latter two support the ESONET specifications, the first two provide own O&M formats.

The ESONET SOS client provides a configuration interface, which allows to specify an appropriate SOS request as well as to format the client’s HTML output, e.g. a title and description. Some SOS servers provide cryptic parameter definitions in urn format - the client allows to transfer this to a human readable format. Furthermore, the number format and accuracy for measurements can be specified. Any SOS specific issue can be defined, e.g. the offering, feature of interest, temporal range of result set etc.

The configuration is saved and used to specify a SOS request which is periodically sent to the SOS server.

The SOS server’s response is also handled by the SOS client. Here, the configuration as well as the data transferred in O&M format by the server is used to generate an internal data format, which is then used to provide the graphical representation of the measurements of an observatory. To embed the information on the overview page of each ESONET site within the ESONET portal, a small chart is generated for each series of measurements and stored in an observatory specific cache directory. In this directory also the data as well as the SOS response is temporarily saved in order to handle e.g. potential network problems or server failures. Beneath this overview chart, also the last measured value is saved and displayed on the overview page.
In addition to this quick overview on the portal's entry page for each ESONET site, a interactive AJAX based SOS charting tool is also available for the interested user. The chart tool opens after the user clicks on the link ‘chart’. It is written using the ‘flot’ Javascript charting library as well as some Jquery libraries which allow asynchronous loading and displaying the cached SOS data. The tool allows to choose between different offerings and to navigate interactively within the displayed data by zooming and panning, and thus provides some basic data analysis possibilities. The technology is similar to the charting tool used for the data catalogue of the ESONET knowledge base and explained there in further detail. In general, the usability is comparable to the Ifremer SOS client shown above, allowing the user to view and analyse data in a similar environment. For clients which do not support Javascript, we additionally offer an exclusively server based simple charting solution.

Figure 7: Screen shot of the interactive ESONET SOS client interface.

The ESONET Knowledge Base

The ESONET knowledge base is the part of the ESONET data infrastructure which assembles and compiles the data and information provided by the components of the ESONET data infrastructure described above and displays it in a human readable, user friendly way. It allows access to data of any of the components of the ESONET data infrastructure, thus serves as a data portal for the ESONET community.

It offers information from the sensor and observatory information given by the ESONET Sensor Registry, access to archived data from the ESONET data catalogue as well as access to real-time data via SOS interfaces. Furthermore, it allows including useful information from the legal & ethical issues database as well as some more general information on each site, such as the coordinates, descriptions and links to further readings on the main ESONET homepage.

The current version of the ESONET knowledge base is a completely new developed system and has replaced the version we have presented in D19. It is now based on a relatively simple technical architecture which strongly decreased the required maintenance efforts and increased the portability of the ESONET knowledge base. Any information sources used
within the knowledge base are configured in a special XML configuration file, which is interpreted by a set of scripts written in PHP which make extensive use of the standards described above (Fig. 8)

Figure 8: Design study of the ‘Sites View’ page of the knowledge base indicating which standards of the data infrastructure mentioned above are used for each part of the portal. This ‘mock-up’ page was used to implement the ESONET Sites View shown below.

In order to generate the portal pages and additionally can trigger or embed external services such as the SOS clients. The ESONET knowledge base can be found at http://dataportals.pangaea.de/esonet. It consists of three major components, the ESONET ‘Sites View’, which allows a quick overview on the data and information for each ESONET site, the ESONET data catalogue which allows querying within archived data as well as the data analysis and charting tool. Thirdly, the knowledge base allows embedding SOS data and provides a SOS client tool, which was already described in the section above.

**ESONET Sites View**

The ESONET Sites View page is the main entry page of the ESONET knowledge base. It offers a user friendly overview on the most important and recent information and data products for each ESONET site. The site view is navigable through a menu structure which currently lists ESONET sites as well as the ESONET demo sites.

ESONET Sites View provides an overview on the position of the site, as well as some background information and basic metadata such as coordinates etc. on this ESONET site. The user can immediately access a selection of archived datasets related to each site from here or use the ESONET data analysis tool to view this data.

Information on datasets is given as data citations indicating the author, year and title of each listed dataset. Clicking on these data sets will direct the user to the responsible data centre where the data set is hosted. Charting options are indicated by a small chart icon behind the
citation. Additionally, the geographical position of these datasets is indicated in a map. Google Maps was used to draw the map, the position and description of each site or dataset respectively was transferred using a KML file generated by the knowledge base.

Most recent additions to a ESONET data archive can be tracked by users by means of a rss feed which is available for subscription if the data archive supports this feature. The knowledge base also makes use of a potentially available SOS server. It displays a preview chart of selected real-time data retrieved by the SOS, a click on this image will open the ESONET SOS client which allows further analysis. Additionally, text links to further pages which offer full access on both, archived and life data (SOS client) are available at this page as well as other subscription options to data news feeds or alerts feed if available. Also, links and overview information on Sensor Registry data as well as the ESONET Legal & Ethical Issues Database are shown at the Sites View.

Figure 9: Example for a ‘Sites View’ page of the knowledge. The screenshot shows the overview for the Arctic Ocean Site for which no cabled observatory providing a SOS is available.

**ESONET data catalogue**

According to the ESONET data management plan, we have set up a data portal providing data sets available in the ESONET data catalogue. Its flexible and distributed architecture allows to include metadata offered by several data centres as long as they support commonly used exchange and metadata formats. Currently the ESONET data catalogue contains data archived by the WDC-MARE, Ifremer, INGV and HCMR data via SeaDataNet. Archived data sets cover a broad range of geosciences and include data such as composition of sediments, chemical characterization of sea water and biological information, depending on the scientific focus of ESONET’s observatories. The majority of data within the data catalogue is available as Open Access data; however for some data sets a moratorium protects the IPR of the responsible scientists as long as detailed investigations are carried out.
The data portal provides two different search tools: a ‘Simple Search’ and an ‘Advanced Search’. Whereas the simple search represents a simple keyword based search engine similar to common internet search engines such as Google, the advanced search has additional functionalities. It allows narrowing a search by using different categories and can be used to formulate more complex queries by combining these search terms. These queries can be combined with geographical and a date/time related query parameters.

The technical platform for the integration of data archives as well as for the public data catalogue is panFMP10 (PANGAEA Framework for Metadata Portals). PanFMP is a generic and flexible framework for building geoscientific metadata portals independent of content standards for metadata and protocols (Schindler & Diepenbroek, 2008). Data providers can be harvested with commonly used protocols (e.g., OAI-PMH) and all metadata standards mentioned above are supported. It uses exclusively Open Source Software and consequently, PanFMP was recently released as Open Source and can be downloaded via SourceForge.

**Simple search** offers a user friendly interface to search for ESONET relevant data via a simple ‘Google-like’ search tool offering a single entry field to define queries for full text search operations on the ESONET metadata catalogue (e.g. type “Guadiana Estuary” as in Google the formats “Guadiana Estuary”, Guadiana-Estuary and Guadiana_Estuary will search for the whole phrase).

**Figure 10:** The simple search window of the ESONET data catalogue. Note the drop down box which assists the user by suggesting several search terms.

The user is assisted by a simple ‘search term suggest wizard’ whose appearance (term list) changes after each entered character. This wizard offers a selection of search terms which definitely can be found in the metadata catalogue.

Search results are presented similar to the commonly used major internet search engines and are shown below simple search. Each data set is listed with a minimum set of metadata such as author, data set title and parameters.

---

10 [http://www.panfmp.org](http://www.panfmp.org)
The displayed data set title is an active hyperlink which leads the user to the data page. This link is given by the responsible data provider. In most cases this is a page containing the complete set of metadata and a link to the data itself.

By clicking on the ‘Advanced Search’ link a new page opens which contains the ESONET data portal advanced search.

The advanced search offers a variety of additional search criteria such as geographical and temporal coverage selection as well as the possibility to search on distinct metadata fields such as parameters or data set authors. Each of these fields additionally provides its own search term suggest wizard which proposes available - field specific - terms after each keystroke.

As the space below the advanced search is limited, results are displayed on a simple search page. However the advanced search can be activated again by clicking on the link mentioned above.

![Advanced Search Window](image)

The geographical search can be carried out in two ways. By typing the geographic coverage (max/min latitude and max/min longitude) into the related fields and click on 'Search', or by clicking in the wind rose icon which opens a map in which the use can select the appropriate area (bounding box). A temporal search restriction can also be defined in the advanced search window by using the calendar tools or by entering the start and end dates in the appropriate form fields.

The search results are listed data set by data set, 10 data sets per web page. For a first overview, each data set is characterized by a data set title (‘citation’), a reference and some information on the size of the data set. By clicking on the data set title the complete metainformation is shown and the data set can be downloaded or visualized on the web page. Additionally the portal provides a charting tool which currently supports data delivered by the WDC-MARE only. Each data which can be visualised by this tool is indicated in the result list by a small chart icon.
Data analysis and charting

As mentioned previously, each data set delivered in an appropriate data format (currently PANGAEA ASCII as well as ESONET O&M) can be visualized by using the new ESONET data analysis and charting tool. This tool uses the flot Javascript library to visualize the data in a user-friendly way. The tool is started after clicking the chart icon which is shown immediately after each data set’s citation either in the ESONET data catalogue result list or in the ESONET site view data overview list. An asynchronous AJAX data loading component (based on Jquery) transfers the data into a JSON format in the background which is then handled and visualized by the flot charting script. In general, the data analysis and charting tool is completely based on Open Source software libraries. The data analysis and charting tool is used for both, displaying archived data as well as for ‘real-time’ data delivered by an observatory’s SOS server.

![Figure 12: A graphical visualisation of a dataset from the Hausgarten observatory](image)

The tool offers to interactively navigate within a chosen data series set by simply clicking and dragging the new range within the chart. This range appears in a pale red color and the tool zooms into this new range and refreshes the chart. The new selection is now indicated in red color within a small overview chart at the bottom of the main chart. Also this overview chart can be used to navigate within the data.

The type of measurements, the parameters are generally displayed against the Y-axis, whereas the X-axis shows the temporal or spatial extent of a data series. Both, X and Y axis parameters are displayed at the left side of the charting tool. The chart is redrawn after clicking on one of the axis selection lists. Single data items can be viewed and analysed by simply moving the mouse pointer within the chart.

Additionally, the type of the chart can be changed by choosing a different chart type in the selection list beyond the chart. Currently we offer line as well as point chart types for ESONET data visualisations.
Summary
The data infrastructure for ESONET is being designed as a distributed system. Both observatory data and archiving services are already provided by several data centres. Therefore, the main challenge is to provide an architecture based on international standards to implement data management policies and work flows. ESONET has a developing online knowledge base for general information about observatory systems and how project-specific data management fits into larger contexts. Core standards include the Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) suite of standards, namely the OGC standards SensorML, Sensor Registry, Catalogue Service for Web (CS-W), Sensor Observation Service (SOS) and Observations and Measurements (O&M) (Fig. 13). OGC SensorML is an eXtensible Markup Language (XML) for describing sensor systems and processes. A core SensorML profile is being developed in cooperation with EuroSITES and OceanSITES to assure international compatibility of sensor descriptions. SensorML files can be stored in the Sensor Registry, the main catalogue of European ocean observatory sensors. The Sensor Registry consists of a web-based entry module which stores the SensorML files in a native XML database. On top of this database, an OGC CS-W interface provides the standardised methodology to publish and access the Sensor Registry. Access to real-time observatory data is then provided by the SOS.
The SOS service harvests data for automated data flow from observatory nodes to a selected data archive. OGC standards are used to generate International Standardisation Organization (ISO) 19139 compliant metadata descriptions for the harvested data. O&M and ISO data is then uploaded to an appropriate data centre and ingested. Suitable data archives provide data curatorial procedures including quality control for manual or semi-automatic upload of data to the archive. Each data archive further provides XML metadata files on each data set in either ISO 19139 or GCMD DIF format and offers metadata exchange via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface. As quality control and other post-acquisition processing are applied to data, these data can be catalogued alongside their raw counterparts.
The knowledge base uses the above standards to data from observatories via the internet. For example, it uses OAI-PMH to harvest metadata from contributing archives and uses SOS to retrieve the latest data from observatories. The technical platform managing the metadata catalogue of contributing data archives is panFMP (PANGAEA Framework for Metadata Portals, Schindler and Diepenbroek 2008). The knowledge base provides access to sensor metadata via the Sensor Registry and offers access to real-time and archived data delivered by the mentioned SOS servers, as well as access to third party data centres which support the standards described above. It additionally provides interactive data analysis and charting tools.