SEVENTH FRAMEWORK PROGRAMME

ICT PPP

Future Internet

ENVIROFI

The Environmental Observation Web and its Service Applications within the Future Internet

FP7-284898

Collaborative project

D5.3 ENVIROFI data and meta-information specifications

University of Southampton IT Innovation Centre

Deliverable due date: 31/08/2012

Actual submission date: 07/09/2012

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 284898

FUTURE INTERNET PPP

SEVENTH FRAMEWORK PROGRAMME

European Commission
Information Society and Media

Document Control Page

Title	D5.3 ENVIROFI data and meta-information specifications
Creator	IT Innovation Centre
Description	This document outlines the ENVIROFI data and meta-information specifications. The approach we have adopted is to first analyse the domain specific data models, vocabularies, ontologies and metadata found in the ENVIROFI pilots. We then review the available metadata standards in the wider geospatial and environmental domain, highlighting various initiatives that are trying to bring together standards in the geospatial area. We bring this analysis together by outlining a clear ENVIROFI approach to integrating heterogeneous data sources, suggesting a methodology based on acknowledgment of the heterogeneous nature of individual environmental domains and use of access brokers to mediate integration between these domains. We conclude with a set of clear recommendations for the ENVIROFI project.
Publisher	ENVIROFI Consortium
Contributors	Stuart E. Middleton, Stefano Modafferi, Zoheir Sabeur (IT-INNOV) Sven Schade, Carlos Granell, Andrea Perego (EC-JRC) Paolo Mazzetti (CNR) Egly Maria (AIT) Hylke van der Schaaf (IOSB) Jasmin Pielorz (UBIMET) Mike Kobernus (NILU)
Creation date	20/07/2012
Type	Text
Language	en-GB
Rights	copyright “ENVIROFI Consortium”
Audience	☐ internal ☒ public ☐ restricted
Review status	☐ Draft ☒ WP leader accepted ☒ Technical Manager accepted ☒ Coordinator accepted
Action requested	☐ to be revised by Partners ☐ for approval by the WP leader ☐ for approval by the Technical Committee ☐ for approval by the Project Coordinator
Requested deadline	

Revision history

Version	Date	Modified by	Comments
v0.1	20/07/2012	Stuart E. Middleton (ITINNOV)	Initial table of contents and partner contributions
v0.2.1	24/07/2012	Zoheir Sabeur, Stuart E. Middleton (ITINNOV)	Further refinement on ToC and partner contributions
v0.2.x	06/08/2012	Sven Schade, Carlos Granell, Andrea Perego (EC-JRC) Paolo Mazzetti (CNR) Egly Maria (AIT)	Comments on ToC and content suggestions/offers for sections 1.4, 4, 5 and 6 Content added for JRC assigned parts of sections 1.4, 4.1 and first draft provided for section 5 Minor editing of sections 1.4 and 4.1. Content added for JRC assigned parts of sections 4.2 and revised draft provided for section 5 WP1 and WP2 data source sections
v0.3	21/08/2012	Stuart E. Middleton (ITINNOV)	Collated draft and revised section 5 structure after JRC discussions
v0.4	24/08/2012	Stuart E. Middleton, Stefano Modafferi (ITINNOV) Sven Schade, Carlos Granell, Andrea Perego (EC-JRC) Conor Delaney(Intune/MI)	WP3 sections, section 5 input Minor editing of sections 1.4 and 4, new content proposal for section 5, re-structuring of section 6, material for appendix B Minor edits to existing sections in 2.3. A added 2 new tables to 2.3
v0.5	28/08/2012	Stuart E. Middleton (ITINNOV) Paolo Mazzetti (CNR) Egly Maria (AIT)	Version for ATOS to QA with a new big data section and input from AIT and CNR. This version needs input from IOSB before it becomes a release candidate.
v0.6	04/09/2012	Jasmin Pielorz (UBIMET)	Updates of content throughout the document
v0.7	05/09/2012	Mike Kobernus (NILU)	Updates of content throughout the document
v0.8	06/09/2012	Hylke van der Schaaf (IOSB)	Updates of content throughout the document
v1.0	07/09/2012	Antonio Oliván, Jose Lorenzo (ATOS)	QA and Final Approval

Copyright © 2012 ENVIROFI Consortium

Table of Contents

Glossary 7
Abbreviations and Acronyms 7
Executive Summary 10
1 Introduction 11
1.1 Overview of domain data sources 11
1.2 Outline of domain vocabulary, ontologies and metadata 11
1.3 Overview of metadata standards in the geospatial and environmental domain 12
2 ENVIROFI data source formal description 14
2.1 WP1 data sources 14
2.2 WP2 data sources 16
2.3 WP3 data sources 17
3 ENVIROFI domain vocabulary, ontologies and metadata 22
3.1 WP1 vocabulary, ontologies and metadata 22
3.2 WP2 vocabulary, ontologies and metadata 23
3.3 WP3 vocabulary, ontologies and metadata 25
4 Data related standards and information models in the geospatial and environ-mental domain 27
4.1 Geospatial standards 27
4.1.1 Baseline standards within the environmental and geospatial communities 27
4.1.2 Geospatial and environmental related standards in other communities 28
4.1.3 Standards focusing on environmental observations 29
4.1.4 Excursion: Pub/Sub in OGC 30
4.2 Geospatial ontologies and information models 31
4.2.1 Generic Models on Location and Geospatial Information 31
4.2.2 Information Models Related to Environment 32
4.2.3 Big data discussions in the geospatial community 36
5 Heterogeneous data sources integration 37
5.1 Mediation and brokering 38
5.2 Cross-domain data fusion and semantic annotation 40
5.3 Mapping domain vocabularies and ontologies 42
6 Recommendations 44
7 Conclusions 46
8 References 47
9 Annexes 51
9.1 Appendix A - Domain data examples 51
9.1.1 Vienna Tree Cadastre 51
9.1.2 Florence Tree Cadastre 52
9.1.3 NILU AirQUIS Air Quality Data 53

Copyright © 2012 ENVIROFI Consortium

9.1.4 ERDDAP sea surface temperature data 54
9.1.5 EUMETSAT sea surface temperature data 56
9.1.6 AIS shipping reports 56
9.1.7 UBIMET SOS Data Service 57
9.1.8 UBIMET WFS 60
9.2 Appendix B - Geo-spatial ontologies and information model examples 61
9.2.1 O&M in INSPIRE 61
9.2.2 O&M in ENVIROFI WP1 on Biodiversity 62
9.2.3 Model for Linked Sensor Data 62

Copyright © 2012 ENVIROFI Consortium

Index of Figures

Figure 1. Mapping semantic layers to structured levels of data fusion 40
Figure 2. Complex property model for O&M as suggested by INSPIRE [56]........ 61
Figure 3. ENVIROFI TaxMeOn Extensions – WP1 (see also ENVIROFI deliverable D1.3.2)........ 62
Figure 4. Suggested Linked Sensor Data Model [63] 63
Figure 5. Suggested extension to the Linked Sensor Data model accounting for aggregates [3] 64

Index of Tables

Table 1. Abbreviations and Acronyms........9
Table 2. Data and meta models for Observation Model - General........ 34
Table 3. Data and meta models for Observation Model - Biodiversity........ 34
Table 4. Data and meta models for Observation Model - Air Quality, Allergens and Meteorology........ 35
Table 5. Data and meta models for Observation Model - Marine 36
Table 6. References 50

Copyright © 2012 ENVIROFI Consortium

Glossary

The glossary of terms used in this deliverable can be found in the public document "ENVIROFI_Glossary.pdf" available at: http://www.envirofi.eu/

Abbreviations and Acronyms

Term	Explanation
ADMS	Asset Description Metadata Schema
BODC	British Oceanographic Data Centre
CDM	Common Data Model
CEN	European Committee for Standardisation
CF	Climate and Forecasting
CIM	Common Information Model
CRS	Coordinate Reference System
DC	Dublin Core
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
EDB	EuroGEOSS Discovery Broker
EML	Ecological Metadata Language
ENVIROFI	The Environmental Observation Web and its Service Applications within the Future Internet
ERDDAP	Environmental Research Division's Data Access Program
EXI	Efficient XML Interchange
FI(-ware)	Future Internet
FI-PPP	Future Internet Public-Private Partnership
FP7	Framework Programme 7
GBIF	Global Biodiversity Information Facility
GCI	GEOSS Common Infrastructure
GEMET	GEneral Multilingual Environmental Thesaurus
GEOSS	Global Earth Observation System of Systems
GMES	Global Monitoring for Environmental and Security
GML	Geography Mark-up Language
HDF	Hierarchical Data Format
HTTP	Hypertext Transfer Protocol
INSPIRE	Infrastructure for Spatial Information in Europe
IoT	Internet of Things

Copyright © 2012 ENVIROFI Consortium

Term	Explanation
ISA	Interoperability Solutions for European Public Administrations
ISO	International Organization for Standardization
JDL	Joint Directors of Laboratories
JSON	JavaScript Object Notation
LOD	Linked Open Data
LTER-Europe	European Long-Term Ecosystem Research Network
MORIS	Monitoring and Research Information System
netCDF	Network Common Data Form
O&M	Observation & Measurement
OGC	Open Geospatial Consortium
OOR	Open Ontology Repository
OPeNDAP	Open-source Project for a Network Data Access Protocol
OWL	Web Ontology Language
PA	Public Administration
POI	Point Of Interest
RDF	Resource Description Framework
REST	REpresentational State Transfer
SDI	Spatial Data Infrastructure
SDMX	Statistical Data and Metadata Exchange
SEIS	Shared Environmental Information System
SensorML	Sensor Mark-up Language
SKOS	Simple Knowledge Organization System
SOA	Service Oriented Architecture
SOAD	Service Oriented Analysis and Design
SOAP	Simple Object Access Protocol
SOCoP	Spatial Ontology Community of Practice
SOS	Sensor Observation Service
SSO	Stimulus Sensor Observation
SWE	Sensor Web Enablement
SWEET	Semantic Web for Earth and Environmental Technology
SWG	Standards Working Group
UncertML	Uncertainty Mark-up Language
URI	Uniform Resource Identifier

Copyright © 2012 ENVIROFI Consortium

Term	Explanation
W3C	World-Wide Web Consortium
WCS	Web Coverage Service
WFS	Web Feature Service
WG	Working Group
WoT	Web of Things
XML	Extensible Mark-up Language

Table 1. Abbreviations and Acronyms

Copyright © 2012 ENVIROFI Consortium

Executive Summary

This document outlines the ENVIROFI data and meta-information specifications. The approach we have adopted is to first analyse the domain specific data models, vocabularies, ontologies and metadata found in the ENVIROFI pilots. This analysis is supported in the appendix with examples of the data sources described. We then review the available metadata standards in the wider geospatial and environmental domain, highlighting various initiatives that are trying to bring together standards in the geospatial area. We bring this analysis together by outlining an ENVIROFI approach to integrating heterogeneous data sources, suggesting a methodology based on acknowledgment of the heterogeneous nature of individual environmental domains and use of access brokers to mediate integration between these domains. We conclude with a set of recommendations for the ENVIROFI project.

The data sources used within ENVIROFI are representative of the wider environmental sector as a whole, with characteristics including a geo-distributed location, heterogeneous data formats and data structures and being based on a variety of domain specific standards and vocabularies. In the biodiversity domain we see observation records for biological assets (e.g. trees and leaves) across wide areas, and use of a variety of ontologies for species labelling. In the air quality domain we see in-situ air quality monitoring stations generating numerous measurement time series, geo-positioned in areas such as cities. In the marine domain we see in-situ buoy sensors measuring time series alongside spatial measurement maps from satellite sources. Taken as a whole it is clear that services within ENVIROFI (and by extrapolation the Future Internet as a whole) must adopt methodologies for representing and integrating both data and metadata which can cope with such diversity of data.

In parallel to domain specific initiatives there has been a lot of work in the geospatial communities on standardization and meta-modelling. We outline a lot of this work in section 4. From our analysis it is clear that the Future Internet will include a network of diversely modelled and formatted pieces of information and we will heavily depend on mediators and brokers in order to make sense of it.

In line with this approach ENVIROFI suggests the following approach to the integration of heterogeneous data sources:

a) Encode domain specific data models using one or more open data formats.

b) Setup 'access brokers' for each domain data model to OGC SWE.

c) Provide transformation services supporting OGC SWE.

d) Provide data fusion services which can fuse data models from two or more domains.

e) Setup 'access brokers' to transform data encoded in one domain data model to another.

f) Chain 'access brokers' to map one domain to another via an intermediate domain.

Copyright © 2012 ENVIROFI Consortium

1 Introduction

This document outlines the ENVIROFI data and meta-information specifications. The approach we have adopted is to first analyse the domain specific data models, vocabularies, ontologies and metadata found in the ENVIROFI pilots. This analysis is supported in the appendix with examples of the data sources described. We then review the available metadata standards in the wider geospatial and environmental domain, highlighting various initiatives that are trying to bring together standards in the geospatial area. We bring this analysis together by outlining an ENVIROFI approach to integrating heterogeneous data sources, suggesting a methodology based on acknowledgment of the heterogeneous nature of individual environmental domains and use of access brokers to mediate integration between these domains. We conclude with a set of recommendations for the ENVIROFI project.

1.1 Overview of domain data sources

In ENVIROFI we have three domain specific pilots, focussing on the biodiversity domain, air quality domain and marine domain. Each of these domains has a variety of data sources available, and the pilots make use of a number of these. Taken as a whole these data sources demonstrate geo-distribution, heterogeneity and a diverse range of domain specific formats and standards.

In the biodiversity domain we are using observation archives of biodiversity assets such as trees and leaves. We have datasets published for the Vienna and Florence regions outlining in detail individual trees, with geo-location and species information. For leaf recognition we have access to a number of image archives from previously published works and will in WP1 gather more images from crowd sourcing.

In the air quality domain we have access to in-situ sensor time series measurements with details of air quality indexes in Oslo and data for Rundle Island in Australia.

In the marine domain we have a combination of in-situ buoy sensors around the Irish coastline and spatial measurement maps from satellite sources (covering Ireland but also the rest of the world). In addition to water measurements we have ship tracking from AIS sensors, providing location tracks for shipping around the world.

1.2 Outline of domain vocabulary, ontologies and metadata

To support each ENVIROFI data source domain we see a wide variety of domain specific standards for data formats, measurement vocabulary and schema/ontologies. Taken as a whole there is little agreement on standards beyond generic open formats (e.g. XML, RDF, netCDF) between domains.

In the biodiversity domain there is a strong use of RDF for ontology representation, using the W3C Simple Knowledge Organization System (SKOS). SPARQL queries are used to retrieve concepts from triple stores. The standards follow geospatial initiatives such as GEMET, GEOSS and INSPIRE but lack a strong community-wide approach to defining domain vocabularies (e.g. species labels).

Within the Air Quality and Meteorology domains, there are no standardised ontologies, however, within the project we followed geospatial initiatives such as GEMET, GEOSS and INSPIRE where ever possible. In order to develop a consistent approach within the Air Quality domain, NILU has been participating in the voluntary group, GEO AQ COP. The GEO AQ CoP is a self-organized voluntary group that fosters the application of Earth observations to air quality management and science. Its participants and its main beneficiaries are members of national and international science teams, data portals and decision support activities. CoP activities include sharing tools and best practices and facilitation of standards-based networking of air quality data systems using GEOSS data sharing principles.

In the marine domain there is a combination of well standardized formats and metadata conventions (e.g. in the satellite field) and proprietary vocabularies less formally defined (e.g. ERDDAP support for IooS). Activities are on-going in each domain to harmonize standards, but the current situation of heterogeneity is expected to continue for many years to come.

Copyright © 2012 ENVIROFI Consortium

1.3 Overview of metadata standards in the geospatial and environmental domain

Metadata (here, data about data, because data about services is out of the scope of this document) is a long-lasting topic in geospatial and environmental data management. Before going into any detail, it should be noted that the distinction between both – data and metadata – strongly depends on the view and intended use of a particular information resource. For many cases it might be sufficient to capture the properties of an object as a simple attribute and then document additional information, such as units or measure, creator etc. separately (as metadata). In other cases, it might be important to directly provide information about used sensing devices, measurement methods and related uncertainties together with the measured value, i.e. all of these items are handled as 'primary citizen' (aka data). We will revisit this case also further down, as it is particularly relevant to environmental sciences and related observations.

Detailed overviews of metadata for the geospatial and environmental sciences have been provided previously, for example in [7], [62], [41], or more recently in the data centric view to the Spatial Data Infrastructures (SDI) reference model of CEN/TC287 [4]. In essence, a comprehensive conceptual model for metadata for discovery, evaluation and use has been standardized as ISO 19115 [61]. This standards has been widely accepted and used, for example in the context of the INSPIRE Metadata Regulation [53][54] and INSPIRE Data Specifications. Implementations include numerous community specific extensions (aka profiles) [19], for example in domains such as meteorology and biodiversity. We will visit some examples in Section 4, where we also point to the relations of these developments in the Environmental Usage Area with FI-Ware (as input to WP4 activities from an SE point of view).

Considering environmental observations and measurement in particular, two standards should mentioned: The Observations and Measurement (O&M) data model of OGC and ISO, and OGC's Sensor Mark-up Language (SensorML). Both together provide the mechanisms for encoding sensing information. A brief summary of the latest developments (Version 2.0) is provided below; this also includes recent guidelines and patterns of using the above mentioned standards for particular domains. These guidelines have been provided as part of the data specification work on INSPIRE and complement the relatively generic ISO and OGC standards with blue-prints for community profiling.

One of the central topics of community extensions is the integration of domain specific terminology in form of code lists and links to community taxonomies or ontologies (section 3). In some areas, certain taxonomies have become common community practice, but this cannot be assumed for all. Accordingly, approaches for translations have been investigated and current approaches use SKOS and the OWL Alignment API for representing matches [5].

Regarding the connection of geospatial and environmental data with other domains, Dublin Core (DC) [13] serves as the common denominator [62]. Beyond that, just recently, a vocabulary for representing location has been developed within the framework of the Interoperability Solutions for European Public Administrations (ISA) Programme of the Commission and was lifted to W3C for consideration [49]. Furthermore, the ISA Programme offers a new conceptual model for metadata describing 'assets' including software and data model specifications [23]. Both might also be valuable to describe assets of the Environmental Usage Area to other communities.

These later developments fit directly into the Linked Data movement within and outside the geospatial sector. Here, (HTTP) URI are used for identifying resources, they are further described following standards (most prominently RDF), and these resources are inter-linked. By its generality, the Linked Data phenomenon might set an end to the above mentioned subjective distinction between data and metadata, as well as it removes the borders between environmental data, geographic data and non-geographic data. Linked Data can be seen as the long requested tool augmenting mainstream IT with classical SDI [74].

On the political dimension, apart from the above mentioned works on INSPIRE, at European level, the Shared Environmental Information System (SEIS) is not at a stage that involved metadata standards can be listed, however, it is likely to use INSPIRE as a basic infrastructure, i.e. also to base on INSPIRE metadata as defined in the guidelines for metadata and the diverse data specifications, together with metadata standards from other domains, such as the Statistical Data and Metadata Exchange (SDMX) for statistics [27]. The Global Monitoring for Environment and Security (GMES) suggests diverse approaches, also due to the sectorial distinction between its space and in-situ components, and the developments within numerous FP7 projects. Parts of the later (particularly the myOcean project [36])

Copyright © 2012 ENVIROFI Consortium

apply an open brokering approach, in which data providers remain free in their choice between any particular (standard) implementation, but connectors are provided by external software components. In this way, numerous standards can be supported as metadata inputs and multiple standards can be used as exchange/output format. The same brokering approach has also led to a revolution within GEOSS, where it became part of the GEOSS Common Infrastructure (GCI) [31].

Copyright © 2012 ENVIROFI Consortium

2 ENVIROFI data source formal description

2.1 WP1 data sources

Data source	Global Earth Observation System of Systems
Owner	US Environmental Protection Agency, http://www.epa.gov/geoss/
Access details	OGC WCS, OGC WFS, OPeNDAP
Format	NetCDF, GeoTIFF, GML, SHAPE
Statistics	n/a
Example	n/a

Data source	Global Biodiversity Information Facility
Owner	Global Biodiversity Information Facility, http://www.gbif.org/
Access details	OGC WFS, GBIF REST
Format	GML, SHAPE
Statistics	n/a
Example	n/a

Data source	WorldClim
Owner	WorldClim, http://www.worldclim.org/
Access details	OGC WCS
Format	NetCDF, GeoTIFF
Statistics	n/a
Example	n/a

Data source	Vienna tree cadastre
Owner	City of Vienna
Access details	Open government data, WFS server: http://data.wien.gv.at/katalog/baumkataster.html
Format	GeoJSON format encoded in RESTful WFS interface Structure of data: id: unique object identifier type: ‘feature’ geometry: GeoJSON geometry object

	type: 'point' coordinates: {0: <longitude>, 1: <latitude>} geometry_name: 'shape' properties: BAUMNUMMER: Number of the tree (ID) GEBIET: Geographic location of the tree (textual description) STRASSE: Road where the tree is located ART: Species information PFLANZJAHR: Year when the tree was planted STAMMUMFANG: Circumference of the trunk one meter above ground KRONENDURCHMESSER: Diameter of the crown BAUMHOEHE: Height of the tree
Statistics	~120000 trees registered, spacial coverage: city of Vienna
Example	see section "Appendix - Vienna Tree Cadastre"

Data source	Firenze Trees
Owner	City of Florence
Access details	Open Government Data; http://datigis.comune.fi.it/shp/alberi.zip
Format	ESRI Shapefile Data structure: Tree ID Scientific Species Name Common Species Name Location (coordinates) District within Firenze Trunk circumference
Statistics	~75500 trees registered, spacial coverage: city of Florence
Example	see section "Appendix - Florence Tree Cadastre"

Data source	Labelled leaf image datasets
Owner	Various authors (all freely published) Leaves of Plummers island (2006) [70] Huff & Royer (2005) [6] Flavia (2007) [72] Swedish leaf (2001) [73]

Copyright © 2012 ENVIROFI Consortium

Access details	Freely available from public websites
Format	Typically these come as a set of JPEG images with an Excel spreadsheet for metadata. Each dataset uses its own vocabulary for species labelling.
Statistics	Plummers island 2006 [7,681 images, 249 species] Huff & Royer 2005 [1,422 images, 17 regional datasets] Flavia 2007 [1,907 images, 33 species] Swedish leaf 2001 [1,125 images, 15 species]
Example	n/a

2.2 WP2 data sources

Data source	NILU AirQUIS Data Service
Owner	NILU
Access details	Time Series Data or Air Quality Index can be obtained by a proprietary http request; license key and ids of timeseries must be provided by NILU Url format for timeseries: http://dataservice.luftkvalitet.info/onlinedata/timeserie/{version}/?id={id}&format={format}&from={from}&to={to}&key={key} {version}: Which version of the output to use. Available version are: v2 {format}: Which format to use. Available formats are: xml and json {from} and {to}: What time period to output. DateTime pattern must be yyyyMMddHHmm. Maximum 1 week of data per timeseries. If not specified, the last hour of data for each timeserie will be selected. {id}: Which timeserie to get online data for. Multiple timeseries are possible, seperated by comma. {Key}: License key provided by NILU Url format for Air Quality Index: http://dataservice.luftkvalitet.info/airqualityindex/area/{version}/?area={area}&format={format}&hoursback={hoursback}&key={key} {version}: Which version of the output to use. Available version are: v2 {format}: Which format to use. Available formats are: xml and json {area}: Which area to display. Multiple areas are possible, seperated by comma. {hoursback}: How many hours from now used in calculation {Key}: License key provided by NILU
Format	Available formats are xml and JSON;
Statistics	Test data containing 27 timeseries from 10 stations within the city of Oslo
Example	See section “Appendix - NILU AirQUIS Air Quality Data”

Data source	UBIMET SOS
Owner	UBIMET
Access details	SOS Service providing getCapabilities and GetObservation for meteorological

Copyright © 2012 ENVIROFI Consortium

	parameters of weather stations: http://ispacevm15.researchstudio.at/sensors/ubimet/sosparameters
Format	XML
Statistics	
Example	see section "Appendix - UBIMET SOS Data Service"

Data source	UBIMET WFS
Owner	UBIMET
Access details	Provides an OpenGIS Web Feature Service for meteorological parameter forecasts. Can be used for forecast requests for arbitrary coordinates and timepoints
Format	Available Formats include xml and JSON
Statistics	
Example	see section "Appendix - UBIMET WFS"

2.3 WP3 data sources

Data source	ERDDAP on the Marine Institute (http://erddap.marine.ie/erddap/info/index.html)
Owner	Marine Institute
Access details	HTTP, ERDDAP, Free access. The marine institute offer a HTTP service that wraps 10 different databases.
Format	Below is the list of supported formats that the ERDDAP portal can download data in: [.asc] View OPeNDAP-style comma-separated ASCII text. [.csv] Download a comma-separated ASCII text table (line 1: names; line 2: units; ISO 8601 times). [.csvp] Download a .csv file with line 1: name (units). Times are ISO 8601 strings. [.das] View the data's metadata via an OPeNDAP Dataset Attribute Structure (DAS). [.dds] View the data's structure via an OPeNDAP Dataset Descriptor Structure (DDS). [.dods] OPeNDAP clients use this to download the data in the DODS binary format. [.esriAscii] Download an ESRI ASCII file (for lat lon data only; lon must be all below or all above 180). [.graph] View a Make A Graph web page.

Copyright © 2012 ENVIROFI Consortium

	[.help] View a web page with a description of griddap. [.html] View an OPeNDAP-style HTML Data Access Form. [.htmlTable] View a .html web page with the data in a table. Times are ISO 8601 strings. [.json] View a table-like JSON file (missing value = 'null'; times are ISO 8601 strings). [.mat] Download a MATLAB binary file. [.nc] Download a NetCDF-3 binary file with COARDS/CF/THREDDS metadata. [.ncHeader] View the header (the metadata) for the NetCDF-3 file. [.odvTxt] Download time,latitude,longitude,otherVariables as an ODV Generic Spreadsheet File (.txt). [.tsv] Download a tab-separated ASCII text table (line 1: names; line 2: units; ISO 8601 times). [.tsvp] Download a .tsv file with line 1: name (units). Times are ISO 8601 strings. [.xhtml] View an XHTML (XML) file with the data in a table. Times are ISO 8601 strings. There are the following data sources, each with its own structure definition: **[Irish Marine Institute Connemara Model CONN2D]** http://erddap.marine.ie/erddap/griddap/IMI_CONN_2D.dds **[Irish Marine Institute Connemara Model CONN3D]** http://erddap.marine.ie/erddap/griddap/IMI_CONN_3D.dds **[Irish National Tide Gauge Network]** http://erddap.marine.ie/erddap/tabledap/IrishNationalTideGaugeNetwork.dds **[Irish Wave Buoys]** http://erddap.marine.ie/erddap/tabledap/IWaveBNetwork.dds **[Irish Wave Buoys 30 Min]** http://erddap.marine.ie/erddap/tabledap/IWaveBNetwork30Min.dds **[Irish Weather Buoy Network]** http://erddap.marine.ie/erddap/tabledap/IWBNetwork.dds **[MI Tide Prediction]** http://erddap.marine.ie/erddap/tabledap/IMI-TidePrediction.dds **[Model Monthly Means of North East Atlantic Model Physical Properties]** http://erddap.marine.ie/erddap/wms/IMI_Model_Stats/index.html **[Topography, ETOPO1, 0.0166667 degrees, Global (longitude -180 to 180), (Ice Sheet Surface)]** http://erddap.marine.ie/erddap/griddap/etopo180.dds **[Topography, ETOPO1, 0.0166667 degrees, Global (longitude 0 to 360), (Ice Sheet Surface)]** http://erddap.marine.ie/erddap/griddap/etopo360.dds
Statistics	**[Irish Marine Institute Connemara Model CONN2D]** spatial coverage: lat: 52.95089447171561 to 53.72912172247213 spacing: 0.001772727 (even) long: -10.79851171874996 to -8.89648828124996 spacing: 0.002976563 (even)

	temporal coverage: 2012-07-18 to 2012-08-24 spacing 3h 0m 0s (even) **[Irish Marine Institute Connemara Model CONN3D]** spatial coverage: lat: 52.95089447171561 to 53.72912172247213 spacing: 0.001772727 (even) long: -10.79851171874996 to -8.89648828124996 spacing: 0.002976563 (even) altitude: 1m to 20m spacing 1.0 (even) temporal coverage: 2012-07-18 to 2012-08-24 spacing 3h 0m 0s (even) **[Irish National Tide Gauge Network]** spatial coverage: lat: 51.6496 to 55.3717; long: --9.9034 to -5.9217 temporal coverage: 2007-01-01to currently 13 Stations (21/08/2012) **[Irish Wave Buoys]** spatial coverage: North bound latitude 54.5; West bound longitude -10; East bound longitude -9; South bound latitude 53 temporal coverage: 2008-07-16 to currently 3 Stations (18/06/2012) 1 Station (21/08/2012) **[Irish Wave Buoys 30 Min]** spatial coverage: Long: -10.278 to -9.271; Lat: 53.227 to 54.28 3 Stations (18/06/2012) 1 Station (21/08/2012) **[Irish Weather Buoy Network]** Spatial coverage: North bound latitude 55.7386 ; West bound longitude -10.5541; East bound longitude -8.4026; South bound latitude 53.8535; temporal coverage: 2007-05-04T06 to currently 5 Stations (21/08/2012) **[MI Tide Prediction]** spatial coverage: lat: 51.6496 to 55.3717; long: --9.9034 to -5.9217 temporal coverage: 2012-01-01 to currently 14 Stations (21/08/2012) **[Model Monthly Means]** spatial coverage: lat: 48.0125 to 58.9875 spacing: 0.025 (even)

Copyright © 2012 ENVIROFI Consortium

	long: -17.9875 to -2.0125 spacing: 0.025 (even) temporal coverage: 2012-06-15 to 2012-07-15 spacing : 30 days(even) **[Topography, ETOPO1, 0.0166667 degrees, Global (longitude -180 to 180), (Ice Sheet Surface)]** spatial coverage: lat: -90 to 90 spacing: 0.01666667 (even) long: -180 to -180 spacing: 0.01666667 (even) **[Topography, ETOPO1, 0.0166667 degrees, Global (longitude 0 to 360), (Ice Sheet Surface)]** spatial coverage: lat: -90 to 90 spacing: 0.01666667 (even) long: 0 to 360 spacing: 0.01666667 (even)
Example	see section “Appendix - ERDDAP sea surface temperature data”

Data source	THREDDS Server on the Marine Institute (http://milas.marine.ie/thredds/catalog.html)
Owner	Marine Institute
Access details	HTTP, free access. The Marine Institute offers an HTTP service that published the output from their fluid dynamic ocean models, which they run on a High Performance Computer
Format	[.nc] Download a NetCDF-3 binary file with COARDS/CF/THREDDS metadata
Statistics	Predicts Hydrodynamical, Biolological, Tidal and Wave statistics for the North East Atlantic. Subsets of this data are published on the ERDDAP server (erddap.marine.ie)
Example	n/a netCDF binaries

Data source	Weather UnderGround
Owner	Weather Underground
Access details	A private company that aggregates personal weather station data. The source data in this case is personal weather sensors around Galway bay
Format	And Http feed (internet frame)
Statistics	Live data on temperature, wind speed and wind direction
Example	n/a

Data source	Automatic Identification System (AIS)

Owner	Marinetraffic.com
Access details	AIS feed published by Marinetraffic.com
Format	XML or HTML AIS shipping reports
Statistics	All shipping activity reported in target area (1000's of ships over time)
Example	see section “Appendix - AIS shipping reports”

Data source	EUMETSAT
Owner	EUMETSAT
Access details	Human interaction, Eumetsat data ordering interface application (free use) Regular FTP data downloads of near real-time data can be obtained for a fee
Format	Grib2, http://www.nco.ncep.noaa.gov/pmb/docs/grib2/download/g2clib.documentation GHRSST https://www.ghrsst.org/data/data-descriptions/ using netCDF with CF metadata profile. We focus on product METEOSAT-9 / Marine / Hourly SST: – Latitude – Longitude – Sea Surface Temperature – Correction level
Statistics	Temporal Covering: 21/12/2005 to currently Hourly data Spatial covering: Africa + Europe
Example	see section “Appendix - EUMETSAT sea surface temperature data”

Copyright © 2012 ENVIROFI Consortium

3 ENVIROFI domain vocabulary, ontologies and metadata

3.1 WP1 vocabulary, ontologies and metadata

Metadata source	GEMET
Owner	EEA
Type	Thesaurus
Access details	SPARQL
Format	SKOS
Statistics	n/a

Metadata source	GEOSS EO Vocabulary
Owner	GEOSS
Type	Thesaurus
Access details	SPARQL
Format	SKOS
Statistics	n/a

Metadata source	EuroGEOSS Drought Vocabulary
Owner	EuroGEOSS
Type	Thesaurus
Access details	SPARQL
Format	SKOS
Statistics	n/a

Metadata source	INSPIRE Feature Concept Dictionary and Glossary
Owner	EEA
Type	Thesaurus
Access details	SPARQL
Format	SKOS
Statistics	N/A

Copyright © 2012 ENVIROFI Consortium

Metadata source	GEOSS Societal Benefit Areas
Owner	GEOSS
Type	Thesaurus
Access details	SPARQL
Format	SKOS
Statistics	N/A

3.2 WP2 vocabulary, ontologies and metadata

NILU is currently participating on the development of an Air Quality Ontology, as part of the GEO Group on Earth Observations, Community of Practice (COP). The GEO AQ CoP is a self-organized voluntary group that fosters the application of Earth observations to air quality management and science. Its participants and its main beneficiaries are members of national and international science teams, data portals and decision support activities. CoP activities include sharing tools and best practices and facilitation of standards-based networking of air quality data systems using GEOSS data sharing principles.

NILU's participation will help to achieve the following:

1. Review the international metadata standards in terms of their applicability to air quality and atmospheric composition needs.
2. Establish a community consensus on the metadata structure and terminology for air quality and atmospheric composition.
3. Identify interconnections among different geographic and thematic areas from emissions to observations and model results on atmospheric composition.
4. Provide recommendations on possible ways to implement the consensus metadata structure and terminology.

Once a standardised ontology has been developed, this will be applied to the ENVIROFI project.

Metadata source	WCS 2.0 Standard
Owner	Open Geospatial Consortium (OGC)
Type	IS
Access details	http://www.opengeospatial.org/standards/wcs/
Format	Available online, in print and various electronic media: Document # 09-110r3
Statistics	N/A

Copyright © 2012 ENVIROFI Consortium

Metadata source	WFS 2.0 Standard
Owner	Open Geospatial Consortium (OGC)
Type	IS
Access details	http://www.opengeospatial.org/standards/wfs/
Format	Available online, in print and various electronic media: Document: 09-025r1
Statistics	N/A

Metadata source	WMS 2.0 Standard
Owner	Open Geospatial Consortium (OGC)
Type	IS
Access details	http://www.opengeospatial.org/standards/wms/
Format	Available online, in print and various electronic media: Document # 06-042
Statistics	N/A

Metadata source	SWEET - Semantic Web for Earth and Environmental Terminology
Owner	Propulsion Laboratory of California Institute of Technology
Type	Ontology
Access details	http://sweet.jpl.nasa.gov/
Format	OWL ontology language
Statistics	6000 concepts

Metadata source	ENVIROFI Glossary
Owner	ENVIROFI
Type	Table
Access details	http://www.envirofi.eu/Downloads/Glossary/tabid/4867/Default.aspx
Format	Available online and in print
Statistics	N/A

Copyright © 2012 ENVIROFI Consortium

3.3 WP3 vocabulary, ontologies and metadata

Metadata source	IooS category
Owner	Integrated Ocean Observing System http://www.ioos.gov/
Type	Proprietary work in progress knowledge-base (http://mmisw.org/ont/ioos) used by ERDDAP
Access details	PDF document on web http://www.ioos.gov/library/us_ioos_blueprint_ver1.pdf
Format	Documented vocabulary detailed in a PDF document and used within ERDDAP software
Statistics	26 core variables

Metadata source	EUMETSAT vocabulary
Owner	EUMETSAT
Type	PDF documents describing the measurement concepts
Access details	Documents downloaded free from website
Format	Documented vocabulary detailed in a PDF document and used within ERDDAP software
Statistics	481 products available. In ENVIROFI we are focussing on the METEOSAT-9 / Marine / Hourly SST

Metadata source	netCDF-CF Climate and Forecast (CF) Metadata Conventions
Owner	Community project whose originators were Brian Eaton (NCAR), Jonathan Gregory (UK Met Office), Bob Drach (LLNL), Karl Taylor (LLNL) and Steve Hankin (NOAA)
Type	Documented standard (PDF, HTML) describing a metadata structure for netCDF This is not a vocabulary definition - it is a structure definition with conventions on handling coordinates, time and units.
Access details	Free, Version 1.6, 5 December, 2011 http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.6/cf-conventions.html
Format	Documented vocabulary detailed in a PDF & HTML document Concept name, unit, ancillary variables, data ranges and guidelines on dimensions and coordinates.
Statistics	The CF conventions define attributes which enable the description of data properties that are outside the scope of the COARDS conventions. Attributes include: – Quantities – Coordinates – Dimensions

Copyright © 2012 ENVIROFI Consortium

	– Intervals and cells – Climatological statistics – Data compression for variables with missing values

Copyright © 2012 ENVIROFI Consortium

4 Data related standards and information models in the geospatial and environmental domain

In line with the introduction (Section 1.4) and the on-going work in WP4 we now detail the already available standards and on-going activities in the Environmental Usage Area. Due to the above indicated richness of the domain, we focus particularly on environmental observations, i.e. modelling attributes also together with the processes that created them, and particularly relate to the thematic domains of ENVIROFI. It is not our aim to provide a complete (or near to complete) overview of all available standards relating to data and metadata. For such, we recommend visiting the forum of the GIGAS project [41] or the recent work on the CEN/TC287 SDI Reference Model [4]. More information about the latter, which itself is a resource with restricted access, is also available from ENVIROFI WP4 deliverables. Particularly, deliverable D4.2 'Environmental Architecture', where also the information viewpoint for the Environmental Usage Ares is depicted.

4.1 Geospatial standards

The extensive set of geospatial and environmental (open) standards also covers data models and metadata specifications. These have certainly high relevance for the Environmental Usage Area, but also closely relate to the foundation in the FI-Ware chapter on 'Data and Context Management'. Whilst the general requirement for representing location has already been passed to the core platform project, more specifics are detailed below. Part of these, for example the ISA locations vocabulary, might serve further material for discussions at FI-PPP Architecture Board level.

As previously indicated, the main standardization bodies in the geospatial domain are ISO, OGC and CEN. All of these work on wider community standards, some also on domain specific models. Accordingly, we use their standards as a baseline – but by far not as the only source – for our developments. At the same time we also channel the ENVIROFI project results back to these bodies for further improvement of their specifications (see also ENVIROFI WP4 deliverables). In the following, we will provide an overview of the scenery with examples related to ENVIROFI WP1, WP2 and WP3. This will show the foundation used in ENVIROFI, but also sketch the wider picture and possible connections to investigate in the near future. This particularly includes the on-going work on introducing Pub/Sub in the context of OGC, which provides one of the most obvious connections to FI-Ware that has to be further investigated.

4.1.1 Baseline standards within the environmental and geospatial communities

The central conceptual data model for geospatial information is defined by the ISO and OGC General Feature Model, which has been further extended in the scope of the ORCHESTRA and SANY projects (see also ENVIROFI deliverable D4.2, section 4.2). Additional elements are then detailed in the ISO Spatial Schema, ISO Temporal Schema and the ISO Rules for Application Schemas. The most widely used encoding standard applies XML-Schema following the latest version of the Geography Mark-up Language (GML, version 3.2.1). The later mentioned observation-specific standards further extend these.

In respect to metadata for geospatial data sets and series, ISO 19115 (part 1 [61] and 2 [60]) can be seen as the core conceptual model. Intensive user guidelines could for example provide due to the ESDIN project [33]. Speaking about environmental topics, INSPIRE metadata implementing rules and guidelines [53][54][55] provide a community extension and also political backing.

ISO 19115 is widely applied within the domain, and many thematic extensions and profiles exist. The NatureSDI+ project [37], for example provided an extension to include taxonomies in ISO 19115 (see also below), in that particular case, for including species names for a biodiversity scenario similar to those of ENVIROFI WP1. Here, the extensions can be used to connect to the TaxMeOn ontology and species descriptions there. Extensions provide a way to embrace GBIF metadata [34]. Still in the biodiversity domain, the Ecological Metadata Language (EML) [24] provides a complement. Extensions for the European Long-Term Ecosystem Research Network (LTER-Europe) were developed by the

Copyright © 2012 ENVIROFI Consortium

EnvEurope project [16].

NetCDF (Network Common Data Format) is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. Recent standardization efforts have resulted in netCDF endorsements by several standards bodies including NASA ESDS, FGDC, OGC, etc. Since 1988, netCDF user documentation has recommended use of conventions for representing meaning in data and for encouraging interoperability between data providers, application developers, and data users. NetCDF and related conventions are adopted by many different communities, as reported at [43].

A crucial aspect in the description of information is the characterisation and quantification of the uncertainty of the data. This should be as complete and detailed as possible, in particular when data is used for spatial decision making. UncertML is an XML schema for describing uncertain information, which is capable of describing a range of uncertain quantities. Its descriptive capabilities range from summaries, such as simple statistics (e.g. the mean and variance of an observation), to more complex representations such as parametric distributions at each point of a regular grid, or even jointly over the entire grid [69]. The FP7 UncertWeb project [42] is working to encode the UncertML into the Observation & Measurement (O&M) and the netCDF specifications; these activities have conducted to the definition of O&M-U and netCDF-U conventions.

4.1.2 Geospatial and environmental related standards in other communities

In the wider picture, i.e. when trying to connect environmental data with other domains, Dublin Core (DC) [13] is considered as a common denominator. Cross-walks between INSPIRE metadata and SDMX are under investigations, as well connections between INSPIRE metadata and data models (on utility and government services) with the Common Information Model (CIM), a common standard of the energy sector.

Scientific

Moreover, on the scientific side – where data sharing remains a topic for debate, but huge progress has already been achieved – three items should be mentioned:

- OPeNDAP: The Open-source Project for a Network Data Access Protocol is a data transport architecture and protocol widely used by earth scientists, based on HTTP. OPeNDAP includes standards for encapsulating structured data, annotating the data with attributes and adding semantics that describe the data. The protocol is maintained by OPeNDAP.org.
- HDF: the Hierarchical Data Format is a file format for storing scientific data and a software library that provides high-level APIs and a low-level data interface. HDF5 is a general purpose library and file format for storing scientific data. It is a data model, library, and file format for storing and managing data. It supports an unlimited variety of data types, and is designed for flexible and efficient I/O and for high volume and complex data.
- CDM: Unidata's Common Data Model is an abstract data model for scientific datasets. It merges the netCDF, OPeNDAP, and HDF5 data models into a common API for scientific data.

eGovernment

Particularly related to eGovernment, the Asset Description Metadata Schema (ADMS) Working Group [21] was established as an initiative in the framework of the ISA programme [15] complementing the group defining the Location Vocabulary. An ADMS asset (also referred to as "semantic interoperability asset") is a set of terms, vocabularies, code lists, etc. which can be used by Public Administrations (PAs) to describe a 'domain', such as the health or geospatial information sectors. The INSPIRE Data Specifications, for example, can be perceived as a semantic interoperability asset(s). Work on metadata for software (aka software assets) is on-going. This work on general assets is highly relevant for the Environmental Usage Area, because it provides the connection to eEnvironment and more widely eGovernment services as defined within the Digital Agenda for Europe [14].

Copyright © 2012 ENVIROFI Consortium

The purpose of the ADMS WG is twofold: (a) defining a conceptual model for ADMS assets – which should then be mapped to an XML schema and an RDF vocabulary – and (b) setting up a federated infrastructure for sharing, searching and accessing ADMS assets. Basically, the idea is to link and make interoperable metadata repositories hosted by European Member States (e.g., Denmark's Digitalisérdk, Germany's XRepository, UK's Govtalk) in order to enforce metadata re-use and, possibly, harmonisation, among European PAs.

On April, 18th, 2012, the WG released the 'final' version of the ADMS specifications [22], which will be used to describe assets published in the 'first wave' of the ADMS federation. Such specifications include also the definition of a core vocabulary, named RADion (where 'RAD' stands for 'Repository-Asset-Distribution'), which defines super-classes for the terms shared by ADMS and the W3C Data Catalog vocabulary (DCAT) [50]. Starting from the end of April 2012, both ADMS and RADion will be contributed to the Government Linked Data [47] working group of W3C to follow the W3C Recommendation track.

Web 2.0

At the same time, also the Environmental Usage Area is considering Web 2.0 development and the emerging lightweight data models and services. ENVIROFI enablers such as the WP5-SE-MED-1 and WP5-SE-MED-2 support the OpenSearch [39] standard with geo and time extensions for discovery functionalities, as well as a set of lightweight metadata models such as Atom, Geo RSS, etc.

Again considering the connection to FI-Ware, we spotted close relations to the FI-WARE Data Element Structure Model and the FI-WARE Context Element Structure Model (see also ENVIROFI deliverable D4.2, page 29 onwards).

4.1.3 Standards focusing on environmental observations

As mentioned previously and also in a series of other ENVIROFI deliverables, the OGC Sensor Web Enablement (SWE) suite of standards covers the very heart of the environmental Observation Web, i.e. it provides a central building block for most (if not all) developments in ENVIROFI. Notably, this does not mean that the SWE standards are the one and only source to commit to, but that they provide highly relevant information – if not the foundation – for most of the data and metadata related works. In the context especially the SWE Common, O&M and the SensorML specifications should be investigated. All are part of the SWE suite in version 2.0. Whilst O&M 2.0 has been finalized; the work on SensorML 2.0 is still in progress.

In a nutshell, SWE Common provides the basic building blocks for any observation related data, whilst O&M focusses on observation results and SensorML on the sensing devices [2]. Central relevant parts of the joint data model are presented in Annex B of this document and have already been used in the context of ENVIROFI WP1. It should be noted that these data models and also the accompanying Sensor Web services closely relate to developments in the domains of the Internet of Things (IoT) and the Web of Things (WoT). For this reason, investigations on the relation to the FI-Ware chapter on 'Internet of Things Enablement' have been initiated in 2011 and discussions have already been moved to OGC. For example FI-Ware attended the OGC SWE and IoT workshop in November 2011 in Brussels, Belgium (see also ENVIROFI deliverable D6.2.2).

The standards mentioned above, particularly including O&M [68] and SensorML [67], however, still require adaptation and specialization for using them in a particular domain and application. Due to several degrees of freedom, e.g. on the definition of features of interest and their properties, community data models that are using O&M (and SensorML) as a bases, might still diverge drastically. In order to increase the potential for interoperability between different communities, additional procedures have to be specified. The INSPIRE guidelines on the use of O&M [56] provide exactly that. They provide a clear way of using O&M for modelling environmental observations and several of the INSPIRE Data Specifications provide best practices on their usage. Some examples are provided in Annex B of this document.

Finally, it should be noted that the same data model (i.e. O&M data model) can be encoded in different data formats due to specific purposes or restrictions. For instance, desktop-based and mobile-based

Copyright © 2012 ENVIROFI Consortium

applications may have different requirements in terms of data formats supported and performance on data exchanged. To this respect, we have recently conducted a performance analysis of exchanging SWE O&M datasets in mobile applications to consume and produce observational data [1] This experiment was aimed to analyse to what extend the performance problems related to transmitting and processing potentially large messages encoded using XML (default encoding for SWE O&M) can be alleviated by using alternative uncompressed and compressed formats such as JSON and EXI (Efficient XML Interchange, http://www.w3.org/TR/exi, introduced earlier in Section 4.1.2.

In summary, the analysis results suggested that using EXI with compression (EXI-C) greatly reduced the size of exchanged messages, but added a high overhead to processing times in the client side, which in the case of mobile devices may lead to serious delays. Nevertheless, it can be an appealing alternative if data is exchanged over very slow or unreliable communication links. Under certain conditions, EXI showed a very good trade-off between size reduction and processing times, even when it does not use compression which implies less energy consumption. The disadvantage of using EXI is that it does not reduce the size of observation blocks (according to the O&M data model), which is a major drawback because exchanged data is expected to be composed of observation values in its majority. This issue might be solved by using a different way to encode these values where strings, timestamps and measurement values were not mixed. Additionally, a pragmatic trade-off would be to offer different data formats according to particular situations. For instance, using EXI for capabilities data files (which contain metadata of the server such as name, keywords, information about provider, and list of available observation offerings), sensor descriptions data files (which contain descriptions of procedures as defined by the SensorML specification), and small observation files, whereas EXI-C may be suitable when a large set of observation values must be exchanged or when information is transmitted over very slow or unreliable communication links.

The study also highlighted the role of JSON, which provides a size reduction of 50% on average in files with low-medium content density —understood as the percentage of an XML document that is 'actual data' (e.g., attribute and element values), in contraposition to the portion that is 'structure' (e.g., namespace information, tags, etc.). In the case of very high content density files, the reduction rate is far less important as the content density value is a lower bound for this value. Nevertheless, because JSON presents faster parsing times and it can be seamlessly integrated in Web-based applications using JavaScript, its use in these applications could bring benefits to SWE-based applications when they do not handle large volumes of observations or network bandwidth is not an issue. This is for instance the case for some ENVIROFI pilots and SEs such as the Vienna Tree application.

4.1.4 Excursion: Pub/Sub in OGC

A recent activity (including a data model) of OGC should be highlighted in this setting, too: the Publish/Subscribe (Pub/Sub) SWG. Generally the SWG aims at building a general Publish/Subscribe standard that can be applied to the whole OGC architecture. This means that it includes but goes beyond the SWE framework. The basic idea behind the Pub/Sub activities is to have a common Publish/Subscribe interface that can be coupled as an extension to every OGC service: It could enhance the Web Map Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS) with Pub/Sub capabilities but also the Sensor Observation Service (SOS) – which is part of the SWE suite of OGC standards. This work is important to ENVIROFI, because FI-Ware is proposing Pub/Sub in the chapter on 'Data and Context Management'.

For this purpose the SWG has first worked on a general, abstract Pub/Sub model. This model is independent of any binding (e.g. SOAP, REST) and defines the relevant concepts and operations at an abstract level. This specification is still in development but for the core elements there is already a quite good draft available. In addition to this abstract model, two bindings are currently in development: SOAP and REST. However, both of these specifications are at an earlier stage than the abstract Pub/Sub model.

One important aspect is, however, that the Pub/Sub SWG does only focus on the Pub/Sub mechanism but not on filter language for describing which messages a user wants to subscribe to. The Pub/Sub specification does not prescribe a certain filter language such as OGC Filter Encoding. Instead it offers a mechanism to refer to (and perhaps negotiate) the filter language a subscriber may use. A typical approach would be to rely on OGC Filter Encoding as filter language. In addition also the filter/query

Copyright © 2012 ENVIROFI Consortium

criteria of the different OGC service types might be used for setting the filters for subscriptions.

For many practical applications we anticipate a need for a common filter language that takes into account event processing concepts. With OGC Filter Encoding this is not available. Instead EML or another mainstream IT language (although this might need geospatial enhancements) should be considered for having the necessary filter capabilities. However, this would be an element of future work.

4.2 Geospatial ontologies and information models

The number of emerging vocabularies is already vast, but it still grows. This is a challenge when talking about data integration (see also Section 5), as well as it is an asset that shown awareness of the overall need and willingness to establish community specific terminologies. In this section, we first present the most central information models and vocabularies for the geospatial sector, before specializing on concepts related to the environmental sector. Beyond some broad horizontal vocabularies, we particularly focus on artefacts, which relate to ENVIROFI work packages WP1 (biodiversity), WP2 (air quality) and WP3 (marine domain).

4.2.1 Generic Models on Location and Geospatial Information

Most generally, i.e. regarding the concept of location and geospatial information, the following works should be mentioned in addition to INSPIRE and the ISA Location Core Vocabulary (mostly taken from a discussion tread of the W3C LOD mailing list [25]):

- GeoSPARQL: The OGC has completed work on GeoSPARQL [38]. This is favoured by the likes of (UK mapping agency) Ordnance Survey and has been produced primarily by geospatial experts with an interest in Linked Data.
- NeoGeo: A community effort has produced NeoGeo [17]. This is favoured by the likes of (French mapping agency) IGN and has been produced primarily by linked data experts with an interest in geospatial data. The primary difference between GeoSPARQL and NeoGeo is in the way they handle point, line and polygon literals. Both enjoy significant support and implementation experience.
- schema.org: It inherits things like name, URL and description from schema.org/Thing which are at least analogous to things like Geographic Names and Geographic Identifiers. schema.org includes the 'containedIn' relation but not, AFAICT, borders etc. The schema.org location properties seem closely linked with event vocabulary. Classes include Mountain, Body of Water, Continent etc. The current list of proposed extensions to schema.org [51] does not include anything in this space and there is no (visibly active) discussion associated with schema.org and location. The current vocabularies include only basic classes and properties for locations, such as:
  - addresses (a clone of vCard) http://schema.org/PostalAddress
  - lat/long (a clone of WGS84) http://schema.org/GeoCoordinates
  - geoShape (including boc, circle, line & polygon) http://schema.org/GeoShape
- W3C Point of Interest: activities of the Points of Interest WG [45]seems to have stopped so that the March 2012 draft [46] looks like the final common denominator. This just at a time when more and more data is being published, a lot of it related to locations and, well, points of interest. The ideas behind the POI WG remain as important as ever but it seems that a new focus is necessary if that work is to be leveraged effectively.
- Standards bodies: OGC and W3C are both willing to help if required but what actually *is* required? That's what the proposed community group is to find out. Once known, related activities can be organized. Like any membership organization, both W3C and OGC put the wishes of their members first. Both bodies are very willing to work together.

Copyright © 2012 ENVIROFI Consortium

- Community Group around vocabularies for describing Location and Addresses [48]: this is hardly a new idea and the last thing to do is to fall into the XKCD trap [52]. Nevertheless, we have different organizations having similar but separate conversations at the moment, mostly born of different use cases and perspectives. This is normal but I think some sort of coordination could be beneficial. One possible outcome is a standard that is backwards compatible with GeoSPARQL and NeoGeo and that combines aspects of both. The danger there is that this would lead to an over-complex standard that could never be fully implemented - which is about as big a pointless waste of time as can be imagined. However, the two are close and common ground shouldn't be hard to find. At the other extreme is that everyone carries on in in their own way and, well, people can pick and choose. This seems less than ideal to me. If interoperability between data sets is important then we need to make some effort to coordinate. The gaps seem to be around linked-data friendly INSPIRE standards, particularly with respect to addresses, and in handling geometry literals that can be huge (no one is talking about yet another way to define points, lines and polygons).

Discussions on common denominators or mappings between the existing elements are still on-going. It would be a great success if the above mentioned W3C community group becomes a focal point for this active collaboration.

As one additional source for information, the volunteer Spatial Ontology Community (SOCoP) is in the process of populating an Open Ontology Repository (OOR) “sandbox” with geospatial ontologies [26]. More information is available on some of this at their Wiki site [75], which is worth to follow.

4.2.2 Information Models Related to Environment

At least on a European level, INSPIRE provides a clear case on the need and wide use of community terminologies. This includes the establishment of community data models for 34 environmental topics (here called ‘themes’)[20] and as part of that feature definitions [58] code lists [57]and common vocabularies to be used as part of the INSPIRE metadata, such as the GEMET, or in order to fill specific attribute values, for example the categories for different types of the measurement methods as part of the Data Specification on Environmental Monitoring Facilities [18] Instead of listing all vocabularies suggested within INSPIRE and other initiatives – such as GEOSS – we from now on focus on prominent observation related examples.

As many other data models, also the above mentioned community specific applications of the O&M standard use domain depended vocabularies in order to define valid expressions of the data model and to describe their intended meaning. Most prominently, the following ontologies are used:

- SI Units [30] being the official reference point for most measurement models, but not formalized in the strict sense.
- Semantic Web for Earth and Environmental Terminology (SWEET) [28] provided by NASA as probably the most cited terminology for earth and environmental science.
- GEneral Multilingual Environmental Thesaurus (GEMET) [32] provided by the European Environmental Agency in the scope of the European Environment Information and Observation Network (Eionet).
- EARTh Thesaurus provided by CNR [29] as a multidimensional classificatory and semantic model, which terminological content is derived from various multilingual and monolingual sources of controlled environmental terminology, including GEMET and others (see also ENVIROFI WP1 deliverables).
- The BODC (British Oceanographic Data Centre) P01 Parameter Usage vocabulary or CF (Climate and Forecast) Standard Names provide two examples from embedding third party vocabularies into INSPIRE (here from the Data Specification on Oceanographic geographical features).

Beyond efforts on providing terminology for community specific measures, other work focuses on a more ontological approach to model observations as such. These efforts complement the more

Copyright © 2012 ENVIROFI Consortium

practically and software engineering focused models of ISO and OGC with a semantic engineering approach. The W3C SSN ontology is at the very forefront of these developments [44] Especially the Stimulus-Sensor-Observation (SSO) design pattern was developed as a flexible and extensible starting point for sensor ontologies and Linked Data vocabularies, it acknowledges the work on O&M, but revises it and provides full-alignment with an upper-level ontology (DOLCE Ultra-Light in this case). An approach for exposing sensor data following this model has been provided [63] and an extension for aggregated observations has been suggested [3].

As far as quality information model, in addition to the already mentioned UncertWeb activities related to UncertML, it is important to consider the work of the FP7 GeoViQua project [35]. In this project the ISO 19115 metadata model for quality is used as the base for the GeoViQua quality information model; a set of new queryables is also defined to allow a user to search for datasets “by quality”.

Taking these models as starting point, we can extend the domain specific needs as already identified in ENVIROFI deliverable D5.2.2 as specified below. We begin at the level of observations, such that more general models are included implicitly; for example, using the INSPIRE Guidelines on the use of Observations and Measurements implies use of the INSPIRE Generic Conceptual Model [59] and the before mentioned INSPIRE code lists and feature dictionary. First of all, we see the following models essential for all work on environmental observations. We suggest those as the basis for any more specific profiles for particular domains.

Data and meta models for Observation Model - General		
ISO IS	ISO Standard	ISO 19156:2011 Geographic information -- Observations and measurements (aligned with OGC O&M 2.0)
OGC IS	OGC Standard	SWE Common Data Model Encoding Standard
OGC IS	OGC Standard	Sensor Model Language (SensorML)
OGC IS	OGC Discussion Paper	Uncertainty Markup Language (UnCertML) *(if uncertainty information is available/desired)*
INSPIRE GL	INSPIRE Guide-lines	INSPIRE Guidelines on the use of Observations and Meas-urements
BIPM Res	BIPM Resolution	SI Units as defined in the resolution of 2006
NASA Term	NASA Terminology	SWEET ontologies on earth science
Eionet Voc	Eionet Vocabulary	GEMET multilingual thesaurus for the environmental domain
CNR The	CNR-IIA Thesaurus	EARTh Thesaurus
Aalto Ont	Aalto University Ontology	TaxMeOn: Ontology Model for Managing Changing Scientific Names in Time *(if changing of names should be considered in the model)*
W3C Ont	W3C Ontology	Semantic Sensor Network Ontology, and especially the SSO design pattern
netCDF	Network Common Data Form	NetCDF is a set of software libraries and self-describing, ma-chine-independent data formats that support the creation, ac-cess, and sharing of array-oriented scientific data.

Copyright © 2012 ENVIROFI Consortium

Data and meta models for Observation Model - General		
netCDF-U	netCDF Uncertainty	Uncertainty conventions for the network Common Data Form

Table 2. Data and meta models for Observation Model - General

For the biodiversity domain, as for example targeted in ENVIROFI WP1, we see the following items as desirable to re-use – in addition to the more general items presented in the table above. Inputs from WP1 deliverables have been considered, too, but we abstracted from pilot specific models, such as the data model of the Monitoring and Research Information System (MORIS).

Data and meta models for Observation Model - Biodiversity		
INSPIRE DS	INSPIRE Data Specifications	Bio-geographical regions (Annex III)
INSPIRE DS	INSPIRE Data Specifications	Environmental monitoring facilities (Annex III)
INSPIRE DS	INSPIRE Data Specifications	Habitats and biotopes (Annex III)
INSPIRE DS	INSPIRE Data Specifications	Species distribution (Annex III)
GBIF Tax	GBIF Taxonomy	Backbone taxonomy including species categorisations as used by the Global Biodiversity Information Facility
ENVIROFI Ont	ENVIROFI Ontology	TaxMeOn with extensions as specified in WP1 (see also ENVIROFI deliverable D1.3.2)
ENVIROFI KB	ENVIROFI Knowledge Base	Species knowledge base as developed in WP1 (see also ENVIROFI deliverable D1.3.2)

Table 3. Data and meta models for Observation Model - Biodiversity

For the Air Quality, Allergens and Meteorology domain(s), as for example targeted in ENVIROFI WP2, we see the following items as desirable to re-use – in addition to the more general items presented above. Inputs from WP2 deliverables have been considered, too, but we abstracted from pilot specific models, such as the data models of the NILU Web Service or the NAAF Pollen Service.

Data and meta models for Observation Model – Air Quality, Allergens and Meteorology		
Eionet Voc	Eionet Vocabulary	GEMET multilingual thesaurus for the environmental domain
GRIB	Gridded Binary	Data format used to store historical and forecast weather data approved by the WMO
INSPIRE DS	INSPIRE Data Specifications	Atmospheric conditions (Annex III)

Copyright © 2012 ENVIROFI Consortium

Data and meta models for Observation Model – Air Quality, Allergens and Meteorology		
INSPIRE DS	INSPIRE Data Specifications	Meteorological geographical features (Annex III)
INSPIRE GL	INSPIRE Guide-lines	INSPIRE Guidelines on the use of Observations and Meas-urements
ISO IS	ISO Standard	ISO 19156:2011 Geographic information -- Observations and measurements (aligned with OGC O&M 2.0)
NASA Term	NASA Terminology	SWEET ontologies on earth science
netCDF	Network Common Data Form	NetCDF is a set of software libraries and self-describing, ma-chine-independent data formats that support the creation, ac-cess, and sharing of array-oriented scientific data.
netCDF-U	netCDF Uncertainty	Uncertainty conventions for the network Common Data Form
OGC IS	OGC Standard	SWE Common Data Model Encoding Standard
OGC IS	OGC Standard	Sensor Model Language (SensorML)
OGC IS	OGC Discussion Paper	Uncertainty Markup Language (UnCertML) (if uncertainty information is available/desired)
W3C Ont	W3C Ontology	Semantic Sensor Network Ontology, and especially the SSO design pattern

Table 4. Data and meta models for Observation Model - Air Quality, Allergens and Meteorology

For the marine sector, as for example targeted in ENVIROFI WP3, we see the following items as desirable to re-use – in addition to the more general items presented above. Inputs from WP3 deliverables have been considered, too, but we abstracted from pilot specific models, such as the models used internal to the Oceanographic Modelling System of the Irish Marine Institute.

Data and meta models for Observation Model - Marine		
INSPIRE DS	INSPIRE Data Specifications	Hydrography (Annex I)
INSPIRE DS	INSPIRE Data Specifications	Oceanographic geographical features (Annex III)
INSPIRE DS	INSPIRE Data Specifications	Sea regions (Annex III)
INSPIRE DS	INSPIRE Data Specifications	Energy resources (Annex III)
OGC IS	OGC Candidate Encoding Standard	Water Markup Language (WaterML), a further specialization of O&M for in-situ hydrological observations data
ERDDAP STD	ERDDAP Stan-dards	ERDDAP (the Environmental Research Division's Data Access Program) is a data server that gives you a simple, consistent way to download subsets of scientific datasets in common file

Copyright © 2012 ENVIROFI Consortium

Data and meta models for Observation Model - Marine		
		formats and make graphs and maps [12]
netCDF-U	netCDF Uncertainty	Uncertainty conventions for the network Common Data Form
NetCDF-CF	NetCDF Climate and Forecast (CF) Metadata Conven-tion	EUMETSAT binary data with variable names follow the NetCDF Climate and Forecast (CF) Metadata Convention [11]
GRIB-2	GRIdded Binary	EUMETSAT binary data formatted in GRIB format [40]

Table 5. Data and meta models for Observation Model - Marine

4.2.3 Big data discussions in the geospatial community

There has been an on-going discussion about big data in the geospatial community, which is outlined in detail in deliverable D6.2.2 [9]. A summary is repeated below for context.

It is becoming clear that Big Data in GIScience does not solely mean to inject geospatial context into mainstream Information Technology (IT). It encompasses also the means to overcome the 'long tail of dark data', i.e., the means to make the highly specialized and detailed, thus highly valuable, results of environmental surveys and measurement campaigns accessible to a wide range of potential (initially unintended) user communities. There is some consensus in the impossibility of making terminology concise and consistent, and the need to integrate potentially contradicting micro theories. Bottom-up metadata on use cases will be-come just as important as binary metadata, i.e. information on how two data sets might fit/work together.

The Big Data phenomenon also poses many practical questions to geospatial sciences and to the way science is carried out in general [10]. Especially, data citing and open access are two prominent issues to consider. Conferences and journals might have to re-orient towards the credible publication of applications and data sets, where for example the number of downloads and effective re-use creates reputation. Data curation tasks arise naturally, and a missing theory for big data provenance becomes eminent. One the one hand, we require background information in order to access and re-use third party data, but on the other hand, Big Data cannot be addressed using Big Metadata.

It should be noted that the currently arising cyber-infrastructures to a certain sense conflict with the Big Data notion. While the former basically create (large) silos in their own, the latter would re-aggregate all available information. For ENVIROFI we want to find ways to connect silos, such as access brokers, and not simply import data into a new ENVIROFI data silo.

Copyright © 2012 ENVIROFI Consortium

5 Heterogeneous data sources integration

During the phase 1 of the FI-PPP, i.e. within ENVIROFI, various implementations of Specific Enablers (SEs) are running in parallel. These SEs use data models basically around the concept of observation, because the collection, management and processing of observations are basic functionalities developed within ENVIROFI. For this reason, and as also argued above the OGC O&M data model must be the basis from which further ENVIROFI data models are defined.

We see, however, that the specification of a unique observation data model to support all of the SEs is a desirable but unrealistic approach. In practice, although the notion of observation is commonplace, differences may exist between developments. For instance, we have the case of the MDAF development and the set of enablers dealing with leaf recognition algorithms. Essentially, both developments stress the concept of observation in their data models, and even these are adaptations from the O&M data model, but at the same time may incorporate distinct features that make them different. So, in a dynamic, agile and flexible environment such as the development of ENVIROFI SEs, proposing a default, unique data model may not be likely the best option since future needs may require changes and updates not foreseen in earlier phases.

Initially the idea of having a common data model and a set of profiles was proposed in past deliverables (e.g. D5.2.2). This still remains a good approach and should be followed as in the sense of keeping backward compatibility with the OGC O&M data model. However, we propose a shift in philosophy. The initial aim of defining one data model that would support all of the SEs should be relaxed in order to allow for various data models that fit better in the development of SEs and future application. Additionally, to boost the integration between FIWARE and ENVIROFI, we have to ensure a feasible, easy and bidirectional connection between ENVIROFI SEs and FIWARE Generic Enablers (GEs). These two kinds of enablers shall work together in application scenarios, communicating where necessary using the FI-Ware Pub/Sub communication data model (Section 4.1.4). Data model from ENVIROFI thus should be mapped to a form suitable of supporting a Pub/Sub FI-Ware communication layer. Indeed, D5.2.2 (page 15) already pointed to this direction and came up with the definition of mapping between SensorML, O&M and UncertML data models to data models(s) used by IoT services (FI-Ware).

One solution is to facilitate on the one hand the provision of different environmental data models and, on the other, to enable smooth mappings between data models between SEs and GEs, is the use of a brokering approach. An access broker can automate transformation between domain specific data models, the ENVIROFI ‘recommended’ environmental data model (OGC SWE) and the requirements of the FI-Ware Pub/Sub communication layer. In essence, the transformation logic is delegated to the broker component, which decouples enablers in the integration, combination, and transformation operations. In summary below is an outline of the approach we foresee for integration of heterogeneous data sources in ENVIROFI:

a) Encode domain specific data models using one or more open data formats (JSON, RDF, XML etc.). The schema and vocabulary of these data models will use existing well-established domain specific standards. This allows domain data source providers to publish data formatted using their own familiar data models, lowering the 'cost to publish' of environmental data in FI-Ware.

b) For each domain data model an 'access broker' can be setup to transform data encoded using a domain specific data model to data encoded using a FI-Ware recommended data model. In ENVIROFI we recommend the OGC SWE O&M data model as FI-Ware's 'recommended' environmental data model.

c) Provide various transformation services supporting FI-Ware's 'recommended' environmental data model (OGC SWE) to assist with common tasks. Examples include transformation of data from one coordinate reference system to another.

d) Provide advanced data fusion services which can fuse data models from two or more domains. The fused result data would be formatted according to a requested domain data model or the FI-WARE 'recommended' environmental data model (OGC SWE O&M).

e) Access brokers can also be setup to transform data encoded in one domain specific data model

Copyright © 2012 ENVIROFI Consortium

to another domain specific data model (not just to the FI-Ware 'recommended' environmental data model). This may be more efficient or simply easier than mapping to the FI-Ware 'recommended' environmental data model.

f) Chaining of access brokers can be used to map one domain to another via an intermediate domain. This would add scalability to our interoperability approach, reducing the number of access brokers required by using a kind of 'hub and spoke' approach to access brokers.

In the following sections we introduce the architecture and basic capabilities of the required brokering and mediation framework, before detailing (ENVIROFI) best practices on realising the internally required transformations. We will particularly focus on the handling of semantic annotations and vocabulary matching as integral parts of the proposed brokering middle ware.

5.1 Mediation and brokering

Environmental applications need to address an information realm which is highly heterogeneous (both in terms of semantic content and adopted technologies). Since this heterogeneity stems largely from different user requirements of the diverse communities, it is not reducible beyond a certain extent [71].

Therefore it is necessary to adopt a System-of-Systems approach based on a multi-style architecture [8]. The brokering approach as a means to implement multi-style architectures has demonstrated valuable in many situations and it is fully supported by ENVIROFI enablers. Specifically:

The Discovery Broker enabler (WP5-SE-MED-1), based on the EuroGEOSS Discovery Broker (EDB) is the cornerstone of the brokering platform. This component is able to read and mediate among the many standards and specifications used by different scientific communities. By building bridges among the practices of these communities, the broker makes it possible to search, and discover the resources available from heterogeneous sources.

To address the semantic interoperability challenge, EuroGEOSS, in collaboration with FP7 GENESIS project (http://www.genesis-fp7.eu/), prototyped a Semantic Discovery Broker which harnesses the EDB capacity. It is the basis of the Discovery augmentation component enabler (WP5-SE-MED-2), implementing It implements a “third-party discovery augmentation approach”: enhancing discovery capabilities of infrastructures by developing new components that leverage on existing systems and resources to automatically enrich available geospatial resource description with semantic meta-information. In fact, the Discovery augmentation component EuroGEOSS Semantic Discovery Broker is able to use existing discovery (e.g. catalogs and discovery brokers) and semantic services (e.g. controlled vocabularies, ontologies, and gazetteers) in order to provide users with semantics enabled query capabilities – contributing to bridge a gap which is important for multidisciplinary infrastructures.

The EuroGEOSS Access Broker enabler (WP5-SE-MED-3EAB) makes it possible for users to access and use the datasets resulting from their queries according to a common grid environment they have configured by selecting the following common features: Coordinate Reference System (CRS), spatial resolution, spatial extent (e.g. subset), and data encoding format. This feature is crucial to allow effective integration and analysis of data coming form heterogeneous sources. In normal practice, the manipulation to the data necessary ahead of the analysis has to be done by the user. The EABAccess Broker takes this burden away from the user, thus providing a true added value service.

The EAB carries out this task by supplementing, but not supplanting, the access services providing the datasets requested. This is achieved by brokering the necessary transformation requests (those that the access services are not able to accomplish) to external processing services. The EABAccess Broker is designed to publish a set of different interfaces to support heterogeneous client applications.

Below is a list of specifications supported by the discovery mediation enablers in ENVIROFI.

Copyright © 2012 ENVIROFI Consortium

Service Type	Supported Versions	Description
CDI	1.0.3, 1.3, 1.4	Common data index service; in use by the marine community
Deegree	2.2	Deegree catalog service based on CSW 2.0.2 ISO AP
Extended brokering interface	6.0	Extended catalog interface based on Broker 6.0
GBIF		Global Biodiversity Facility services
GENESI DR		GENESI DR project service interface
GeoNetwork	2.2.0, 2.4.1, 2.6	GeoNetwork catalog service based on CSW 2.0.2 ISO AP
OGC CSW	2.0.2 (CORE, AP ISO, ebRIM/CIM, ebRIM/EO)	Catalogue Services for the Web
OGC WCS	1.0, 1.1	Web Coverage Service
OGC WFS	1.0	Web Feature Service
OGC WMS	1.3.0, 1.1.1	Web Map Service
OpenSearch		OpenSearch engines
THREDDS	1.0.1, 1.0.2	Thematic Realtime Environmental Distributed Data Services
GI-cat Extended Interface	7.x, 8.x, 9.x	Broker discovery interface
THREDDS-NCISO	1.0.1, 1.0.2	THREDDS Data Server with ncISO plug-in
ESRI ArcGIS Geoportal catalog service	10	ESRI Catalog (based on CSW/ISO AP 2.0.2)
OAI-PMH	2.0 (ISO 19139, Dublin Core)	Open Archive Initiative
NetCDF-CF	1.4	Climate and Forecast conventions of netCDF
NcML-CF		XML encoding of netCDF metadata with Climate and Forecasst convetions
NcML-OD		XML encoding of netCDF metadata with Observation Dataset convention
GeoRSS	2.0	Geo extension of RSS
GDACS		Global Disaster Alert and Coordination System
DIF		Directory Interchange Format
File System		Locally stored metadata
SITAD		Sistema Informativo Territoriale Ambientale Diffuso
INPE		Web Catalog of INPE (Brazil Space Agency)
HYDRO		CUAHSI inventory and access service

Copyright © 2012 ENVIROFI Consortium

5.2 Cross-domain data fusion and semantic annotation

The JDL data fusion information model [64] has motivated projects such as SANY and TRIDEC to develop prototype knowledge-based service architectures [66] [65] for scalable and efficient data fusion in the environmental domains. These projects separated into distinct semantic layers the results of data fusion at different JDL levels. These semantic layers can be seen in Figure 1.

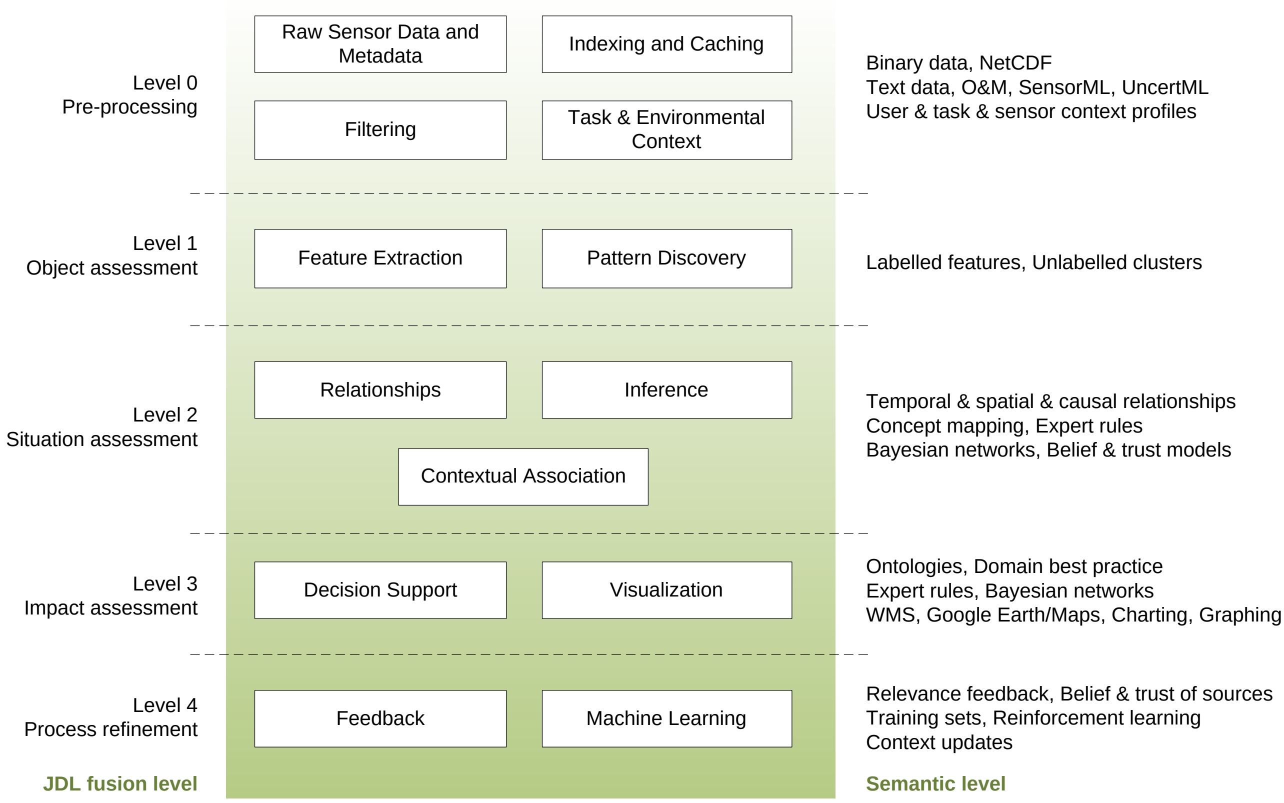

Figure 1. Mapping semantic layers to structured levels of data fusion

Building on these results in ENVIROFI we can address some of the issues of heterogeneous data integration at each of the different levels of data fusion. This provides an approach allowing data fusion to be performed across datasets from heterogeneous domains in a scalable and metadata driven way.

At the pre-processing level (JDL level 0) we can address issues relating to heterogeneous vocabularies used by different domains to label their datasets. Typical activities at this level of fusion are domain specific syntax checking, cleaning and preparation of data ready for work at higher levels of fusion. In ENVIROFI we allow users to specify a target domain for fusion results and use vocabulary mapping techniques (see section 5.3) to map concepts from data sources to the target domain for the result data. The use of access brokers can greatly help this process as they allow access to data sources and transformation of data from unfamiliar domains into a format and vocabulary that is familiar.

In ENVIROFI we have examples of pre-processing where we take data from two WP3 sources: ERDDAP sea surface temperature time series and EUMETSAT satellite sea surface temperature maps. This data is mapped to a preferred target domain vocabulary, ready for subsequent data fusion work at high levels.

At the object assessment level (JDL level 1) we can perform low level feature extraction and pattern discovery on pre-processed data to semantically annotate datasets and data streams. Work at this level includes looking for patterns in time series measurements (e.g. regressing over historical data), image processing (looking for patterns in shape, colour, texture, etc.), natural language processing (applying n-Gram token extraction, grammar parsing, etc.) and spatial map analysis (identifying coherent regions etc.). The objects identified in the pre-processed dataset are used to add semantic annotations to the

Copyright © 2012 ENVIROFI Consortium

dataset. Whilst feature extraction algorithms themselves normally require domain specific training data or configuration the semantic annotations they produce can be abstracted and domain-neutral, allowing for the possibility of cross-domain analysis when working at higher levels of fusion.

An example of object assessment can be seen in ENVIROFI where we are taking unlabelled images of leaves and using a shape-based image processing algorithm to create shape outline signatures that can be used with a training set to label images according to appropriate domain-specific species ontologies. These image labels also allow work at higher levels of fusion to combine labelled image data from regions of interest (e.g. UK, Italy, Austria) with different types of knowledge (e.g. habitat information).

At the situation assessment level (JDL level 2) input data has usually been pre-processed to use a homogeneous structure and vocabulary, allowing more complex data fusion work to intelligently bring together the information into a common frame of reference within a spatial and temporal region of interest. Typically the kind of operations that are performed at this level include temporal, spatial and spatial-temporal interpolation and clustering work. The resulting situation assessment picture is returned using a preferred target domain's vocabulary.

Examples of situation assessment can be seen in ENVIROFI where we are taking EUMETSAT maps and ERDDAP point measurements of sea surface temperature and combining them together using both spatial and temporal interpolation techniques. A fused situation picture is then provide using the target domain's vocabulary.

The combination of data fusion processes operating at different levels of fusion fits well with a broker pattern and can be presented via access broker endpoints, allowing access to fused data in open data formats such as XML and RDF. This can be seen in work on the fusion toolkit mediator and uncertainty annotation services.

Copyright © 2012 ENVIROFI Consortium

5.3 Mapping domain vocabularies and ontologies

The last decade has seen a growing adoption of information systems with a semantic-enabled repository, generally in the form of the so called triple stores. Usually, these systems rely on knowledge base access modules tightly integrated into the system. These modules usually offer the opportunity to directly access and manipulate their content through application programming interfaces (APIs). However, in a service-oriented architecture (SOA) as proposed in ENVIROFI, where the semantic repository would be a service as well, this solution has some limitations. In fact, the implementations of these APIs normally do not support any kind of network protocol or service interface.

To access and manipulate ontologies there are several APIs available. Most of them are for the programming language Java and feature a fine grained and therefore powerful interface. The most popular is still Jena but it currently does not support OWL 2. For full OWL 2 support there is the OWL API [76]. The OWL API is a Java API and reference implementation for creating, manipulating and serializing OWL Ontologies. Starting with version 3 it supports the full OWL 2 standard. It is currently evolving into a kind of de facto standard; for example it is used by the most widespread ontology editor Protégé 4 [77]. All of these APIs are designed for direct integration into a system and do not provide a network abstraction. A brute force approach to design a service interface could be to create a 1:1 mapping from e.g., the OWL API. There are several drawbacks with such an attempt:

- In a SOA the communication should be kept minimal because it is time consuming and slows down the overall system. For example if one would add some instances to the ontology, many operations must be called and therefore much communication takes place.
- The available APIs are mostly designed object-oriented, which often means that if you call a function you get back an object on which you call further functions (e.g., get a concept and then get its object properties). In contrast, service interfaces tend to be not object-oriented and have a flat set of operations. However it could be done by using complex data structures which would be complicated to implement and difficult to use (and again causes a lot of communication).

Instead of using an API there is the possibility to use query languages (QLs) for Semantic Web ontologies. An example of such a language is SPARQL [78]. With its extension *SPARQL Update* it can cover most API functionality. A drawback is that modifications to the ontology which are fairly simple via an API tend to get quite complex if formulated in SPARQL. SPARQL acts on the same level as SQL for relational databases and has the possibility to be used via a network with the Remote SPARQL specification, which compares to JDBC or ODBC for SQL. The architecture suggested in the PESCaDO project tries to combine both approaches of API and query language to limit the drawbacks [79].

A further aspect to be considered when developing a service-based semantic repository is the trade-off between size and reasoning capabilities supported, especially in real time and interactive systems: reasoning performance falls with the size of the knowledge base. Typically, this aspect is tackled by supporting restricted profiles of the used ontology language (e.g., OWL 2 QL), thus limiting the reasoning capabilities of the system.

However, there are situations in which alternative strategies could also be considered. In various applications, it may happen that the inferred content to be produced via reasoning may be limited to a specific user session or to limited content of the knowledge base. When working with domain-specific measurements only those parts of knowledge base need to be considered that are relevant to the domain. In such situations, one can think of extracting from the general data repository only the data which are relevant for the considered user session, to instantiate a user session specific knowledge-base with the relevant data, and to perform reasoning on this smaller set. This approach, referred to as session-based ontology instantiation, allows an efficient exploitation of the reasoning capabilities offered by more expressive variants of the ontology language used (e.g., OWL 2 DL), due to the limited size of the session specific knowledge base.

A further benefit of the session-based ontology instantiation approach is that, due to its limited size, it enables working with an in-memory implementation of the session-specific knowledge base instead of a persistent one, thus favouring a further improvement of the performance of the system.

This architecture is also a good starting point for a federated ontology architecture, where the knowl-

Copyright © 2012 ENVIROFI Consortium

edge is distributed over several services. Indeed, establishing a *master ontology* for a larger community is considered by domain-experts to be infeasible and thus a larger community will have to work with a distributed set of ontologies. A reasoning request on such a federated knowledge base would start with the instantiation of a session knowledge-base containing the relevant parts of the distributed knowledge bases and the actual reasoning would be done on this session instance. One challenge with such a federated architecture is how to align the data from the different knowledge-bases that respond to a federated search request, since a suitable cross-walk is often not available. The available ontologies might use different terminology for the same concepts, contain similar concepts that can not be matched directly, are not equivalent, or even contain conflicting data. How to deal with issues like this is a topic of ongoing research.

Copyright © 2012 ENVIROFI Consortium

6 Recommendations

Based on the discussions above – especially from a WP1-WP2-WP3 point of view, but also acknowledging other on-going works within the geospatial and environmental usage area – we propose the following.

Recommendation 1

Recognize that in the environmental domain there are already several working information sharing systems (international like GEOSS, thematic like GBIF, etc.); therefore never propose to replace them, instead propose to supplement them adding advanced (interoperability) functionalities, according to a System-of-Systems approach

Recommendation 2

Architectural heterogeneity cannot be completely eliminated since it often comes from different user requirements. Suggest the adoption of multi-style architectures for building environmental System-of-Systems. The brokering approach as a means to implement multi-style architectures has demonstrated valuable and it is supported by ENVIROFI enablers.

Recommendation 3

Standardization as a means of interoperability helps to reduce information heterogeneity avoiding the proliferation of specifications. Never impose the use of a particular standard on anybody, but advertise those with which we can already work today.

Implementation note: Always aim at investigating the gabs and then negotiate if one party should change/align its own approach or if mediation should be provided instead.

Implementation note: Use standards and vocabularies identified in the tables in section 4.2 wherever possible (to ensure some degree of interoperability or at least easy adoption following a brokering approach).

Recommendation 4

Continue the investigations on reference frames whenever possible.

Implementation note: Rely on OGC/ISO O&M as a baseline and facilitate available guidelines and community vocabularies.

Implementation note: Further follow the ISA Programme in order to investigate benefits for using their work an asset metadata.

Implementation note: Follow the migration of the location ontology to W3C and the promote integration of this in the Environmental Usage Area.

Implementation note: Actively participate in the W3C community group on 'Location and Addresses'.

Copyright © 2012 ENVIROFI Consortium

Recommendation 5

Further follow and investigate FI-Ware chapters on 'IoT Enablement' and 'Data and Context Management' in order to ensure smooth integration with information exchange between environmental enablers and future GEs

Implementation note: Connect and follow the work of the OGC Pub/Sub working group and the related FI-Ware activities in order to establish common interfaces and data models for future integrations of FI-Ware GEs with SEs on environmental observation. This should include investigations on the possible need for a common filter language that takes into account event processing concepts..

Implementation note: Extend the brokering framework in order to support also Pub/Sub for data exchange. This should facilitate smooth interaction with FI-Ware GEs, particularly related to the chapter on 'Data and Context Management'. In this context, the exact intentions of FI-Ware have to be cross-checked. Currently the 'Data and Context Management' chapter speaks about the Pub/Sub model, a Pub/Sub broker and a context broker. The different roles of these items remain unclear.

Recommendation 6

Further follow and investigate relations to other use case projects in and outside the FI-PPP in order to connect used (metadata) data and information models.

Implementation note: Further try to collaborate with the other Use Case projects in order to investigate possible use of an environmental observation and measurements model for the Future Internet.

Copyright © 2012 ENVIROFI Consortium

7 Conclusions

The data sources used within ENVIROFI are representative of the wider environmental sector as a whole, with characteristics including a geo-distributed location, heterogeneous data formats and data structures and being based on a variety of domain specific standards and vocabularies. In the biodiversity domain we see observation records for biological assets (e.g. trees and leaves) across wide areas, and using a variety of ontologies for species labelling. In the air quality domain we see in-situ air quality monitoring stations generating numerous measurement time series, geo-positioned in areas such as cities. In the marine domain we see in-situ buoy sensors measuring time series alongside spatial measurement maps from satellite sources. Taken as a whole it is clear that services within ENVIROFI (and by extrapolation the Future Internet as a whole) must adopt methodologies for representing and integrating both data and metadata which can cope with diversity of data.

In parallel to domain specific initiatives there has been a lot of work in the geospatial communities on standardization and meta-modelling. We outline a lot of this work in section 4. From our analysis it is clear that the Future Internet will include a network of diversity modelled and formatted pieces of information and we will heavily depend on mediators and brokers in order to make sense of it.

In line with this approach ENVIROFI suggests the following approach to the integration of heterogeneous data sources:

a) Encode domain specific data models using one or more open data formats.

b) Setup 'access brokers' for each domain data model to support FI-WARE's 'recommended' environmental data model (OGC SWE).

c) Provide transformation services supporting FI-WARE's 'recommended' environmental data model (OGC SWE).

d) Provide data fusion services which can fuse data models from two or more domains.

e) Setup 'access brokers' to transform data encoded in one domain data model to another.

f) Chain 'access brokers' to map one domain to another via an intermediate domain.

Copyright © 2012 ENVIROFI Consortium

8 References

References

[1] A. Tamayo, C. Granell, and J. Huerta (2012). Using SWE Standards for Ubiquitous Environmental Sensing: A Performance Analysis. Sensors (in press)

[2] Arne Bröring, Johannes Echterhoff, Simon Jirka, Ingo Simonis, Thomas Everding, Christoph Stasch, Steve Liang and Rob Lemmens (2011). New Generation Sensor Web Enablement. Sensors, 11(3), pp. 2652-2699.

[3] C. Stasch, S. Schade, A. Llaves, K. Janowicz and A. Broering (2011). Aggregating Linked Sensor Data. Semantic Sensor Network Workshop 2011, Bonn, Germany.

[4] CEN/TC287 (2012). Geographic information — Spatial Data Infrastructures — Part 3 : Data centric view.

[5] Cristiano Fugazza, Sören Dupke and Lorenzino Vaccari (2010). Matching SKOS Thesauri for Spatial Data Infrastructures. MTSR 2010: 211-221.

[6] D.L. Royer, P. W. (2005). Correlations of climate and plant ecology to leaf size and shape: potential proxies for the fossil record. American Journal of Botany, 1141-1151.

[7] Douglas Nebert (2004). Developing Spatial Data Infrastructures: The SDI Cookbook. GSDI, 2004

[8] ENVIROFI Deliverable 6.1.1 (2011). Sketch of ENVIROFI Architecture.

[9] ENVIROFI Deliverable 6.2.2 (2012). Report on Community Consultation II

[10] ERCIM special issue on 'e-infrastructures for Big Data: Opportunities and Challenges', April 2012, http://ercim-news.ercim.eu/images/stories/EN89/EN89-web.pdf

[11] http://cf-pcmdi.llnl.gov/

[12] http://coastwatch.pfeg.noaa.gov/erddap/index.html

[13] http://dublincore.org/

[14] http://ec.europa.eu/information_society/digital-agenda/index_en.htm

[15] http://ec.europa.eu/isa/

[16] http://enveurope.geocatalogue.ise.cnr.it/deims/sites/default/files/EnvEurope_MD_Specification_FinalDraft_v2.0.pdf

[17] http://geovocab.org/doc/neogeo/

[18] http://inspire.ec.europa.eu/codeList/ResultAcquisitionSourceValue

[19] http://inspire.ec.europa.eu/index.cfm/pageid/2

[20] http://inspire.ec.europa.eu/index.cfm/pageid/2

[21] http://joinup.ec.europa.eu/asset/adms/home

[22] http://joinup.ec.europa.eu/asset/adms/release/100

[23] http://joinup.ec.europa.eu/asset/core_location/description

[24] http://knb.ecoinformatics.org/software/eml/

FUTURE INTERNET PPP Copyright © 2012 ENVIROFI Consortium

[25] http://lists.w3.org/Archives/Public/public-lod/2012Aug/0009.html

[26] http://oor-01.cim3.net/ontologies

[27] http://sdmx.org/

[28] http://sweet.jpl.nasa.gov/ontology/

[29] http://uta.iia.cnr.it/earth_eng.htm

[30] http://www.bipm.org/en/CGPM/db/11/12/

[31] http://www.earthobservations.org/gci_gci.shtml

[32] http://www.eionet.europa.eu/gemet/

[33] http://www.esdin.eu/project/summary-esdin-project-public-deliverables#metadata

[34] http://www.gbif.org/informatics/discoverymetadata/current-metadata-system/

[35] http://www.geoviqua.org

[36] http://www.myocean.eu.org/

[37] http://www.nature-sdi.eu/

[38] http://www.opengeospatial.org/standards/geosparql

[39] http://www.opensearch.org

[40] http://www.osi-saf.org/biblio/docs/ss1_pmatlsst_1_6.pdf

[41] http://www.thegigasforum.eu/forum/welcome.html

[42] http://www.uncertweb.org

[43] http://www.unidata.ucar.edu/software/netcdf

[44] http://www.w3.org/2005/Incubator/ssn/wiki/Main_Page

[45] http://www.w3.org/2010/POI/

[46] http://www.w3.org/2010/POI/documents/Core/core-20111216.html

[47] http://www.w3.org/2011/gld/

[48] http://www.w3.org/community/groups/proposed/#locadd

[49] http://www.w3.org/QA/2012/06/implementing_adms_and_the_core.html

[50] http://www.w3.org/TR/vocab-dcat/

[51] http://www.w3.org/wiki/WebSchemas/SchemaDotOrgProposals

[52] http://xkcd.com/927/

[53] INSPIRE (2008). INSPIRE Metadata Regulation 03.12.2008.

[54] INSPIRE (2009). Corrigendum to INSPIRE Metadata Regulation 15.12.2009.

[55] INSPIRE (2010).INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Version 1.2) 16.06.2010.

[56] INSPIRE (2011). D2.9 Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE Annex II and III data specification development.

Copyright © 2012 ENVIROFI Consortium

[57] INSPIRE Code List Dictionaries (April 2010) 03.05.2010

[58] INSPIRE Feature Concept Dictionary 19.12.2008

[59] INSPIRE Generic Conceptual Model 31.07.2012

[60] ISO 19115-2. Geographic information -- Metadata -- Part 2: Extensions for imagery and gridded data.

[61] ISO/TC211 (2003). ISO 19115: international metadata standard for geographic information.

[62] Javier Nogueras-Iso, F. Javier Zarazaga-Soria, Pedro R. Muro-Medrano: Geographic information metadata for spatial data infrastructures - resources, interoperability and information retrieval. Springer 2005: I-XXI, 1-263

[63] K. Janowicz, A. Broering, C. Stasch, S. Schade, T. Everding and A. Llaves (2011). A RESTful Proxy and Data Model for Linked Sensor Data. International Journal of Digital Earth.

[64] Lambert, D. A.: A blueprint for higher-level fusion systems. Information Fusion, 10 (1), 6-24 (2009)

[65] Middleton, S.E. (eds.): SANY fusion and modelling architecture, OGC discussion paper, OGC ref 10-001 (2010)

[66] Middleton, S.E. Sabeur, Z.A. "Knowledge-based Service Architecture for Multi-risk Environmental Decision Support Applications", ISESS 2011

[67] OGC (2007). OGC Implementation Specification 07-000: OpenGIS Sensor Model Language (SensorML); Open Geospatial Consortium: Wayland, MA, USA, 2007.

[68] OGC (2011). OGC Implementation Specification 10-025r1: Observations and Measurements - XML Implementation; Open Geospatial Consortium: Wayland, MA, USA, 2007.

[69] Open Geospatial Consortium (2009). Uncertainty Markup Language (UnCertML). OGC Discussion Paper 08-122r2

[70] P. Belhumeur, D. C. (2008). Searching the World's Herbaria: A System for Visual Identification of Plant Species. ECCV, (pp. 116-129).

[71] P. Mazzetti, S. Nativi (2012). Multidisciplinary Interoperability for Earth Observations: Some Architectural Issues. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 5, Issue 3

[72] S. Gang Wu, F. S. (2007). A Leaf Recognition Algorithm for Plant classification Using Probabilistic Neural Network. IEEE 7th International Symposium on Signal Processing and Information Technology. Cario, Egypt.

[73] Söderkvist, O. (2001). Computer Vision Classification of Leaves from Swedish Trees. Linköping Univ.

[74] Sven Schade and Paul Smits (2012). Why Linked Data Should Not Lead to Next Generation SDI. IGARSS 2012, Munich, Germany.

[75] www.socop.org

[76] M. Horridge, S. Bechhofer, "The OWL API: A Java API for working with OWL 2 ontologies," OWLED, volume 529, Proceedings of CEUR Workshop, 2008.

[77] http://protege.stanford.edu/

[78] E. Prud'hommeaux, A. Seaborne, "Sparql query language for rdf," W3C Working Draft, http://www.w3.org/TR/rdf-sparql-query/ , 2006.

Copyright © 2012 ENVIROFI Consortium

[79] J. Moßgraber, M. Rospocher (2012). Ontology Management in a Service-oriented Architecture. 11th International Workshop on Web Semantics and Information Processing.

Table 6. References

Copyright © 2012 ENVIROFI Consortium

9 Annexes

9.1 Appendix A - Domain data examples

9.1.1 Vienna Tree Cadastre

The data of the Viennese tree registry are available in geoJSON format:


```
{
   "type":"FeatureCollection",
   "features":
   [
          {
                 "type":"Feature",
                 "id":"BAUMOGD.5483724",
                 "geometry":
                 {
                        "type":"Point",
                        "coordinates":[
                               16.466546227665713,
                               48.234957805653934
                        ]
                 },
                 "geometry_name":"SHAPE",
                 "properties":
                 {
                        "BAUMNUMMER":"139  ",
                        "GEBIET":"Strassen",
                        "STRASSE":"Hirschstettner Strasse",
                        "ART":"Ulmus sp. (Ulme)",
                        "PFLANZJAHR":1953,
                        "STAMMUMFANG":211,
                        "KRONENDURCHMESSER":7,
                        "BAUMHOEHE":8
                 }
          },
          {
                 "type":"Feature",
                 "id":"BAUMOGD.5483725",
                 "geometry":
                 {
                        "type":"Point",
                        "coordinates":[
                               16.466618464918067,48.23493784876904
                        ]
                 },
                 "geometry_name":"SHAPE",
                 "properties":
                 {
                        "BAUMNUMMER":"140  ",
                        "GEBIET":"Strassen",
                        "STRASSE":"Hirschstettner Strasse",
                        "ART":"Fraxinus excelsior (Gemeine Esche)",
                        "PFLANZJAHR":1998,
                        "STAMMUMFANG":46,
                        "KRONENDURCHMESSER":2,
                        "BAUMHOEHE":7
                 }
          },
          {
                 "type":"Feature",
                 "id":"BAUMOGD.5483726",
                 "geometry":
                 {
                        "type":"Point",
                        "coordinates":
                        [
                               16.39204790504051,
                               48.289325107247066
                        ]
                 },
```

```
                 "geometry_name":"SHAPE",
                 "properties":
                 {
                        "BAUMNUMMER":"218  ",
                        "GEBIET":"Strassen",
                        "STRASSE":"Fritz-Kandl-Gasse",
                        "ART":"Gleditsia triacanthos 'Skyline' (Säulengleditschie)",
                        "PFLANZJAHR":2011,
                        "STAMMUMFANG":19,
                        "KRONENDURCHMESSER":1,
                        "BAUMHOEHE":3
                 }
          }
   ]
}
```


9.1.2 Florence Tree Cadastre

Below is an example ESRI shapefile showing the structure of the Florence tree data:

```
CODSITO    QUARTIERE      SPECIE  NOMECOMUNE     CIRCCM
1  43461   3       Pinus pinea    Pino domestico 211
3  43463   3       Pinus pinea    Pino domestico 187
4  43464   3       Pinus pinea    Pino domestico 131
5  43465   3       Pinus pinea    Pino domestico 114
6  43466   3       Pinus pinea    Pino domestico 181
7  43467   3       Pinus pinea    Pino domestico 184
8  43468   3       Pinus pinea    Pino domestico 190
9  43469   3       Pinus pinea    Pino domestico 157
10 43470   3       Pinus pinea    Pino domestico 198
11 43471   3       Pinus pinea    Pino domestico 146
13 43473   3       Pinus pinea    Pino domestico 178
14 43474   3       Pinus pinea    Pino domestico 158
15 43475   3       Pinus pinea    Pino domestico 171
16 43476   3       Pinus pinea    Pino domestico 175
17 43477   3       Pinus pinea    Pino domestico 158
18 43478   3       Pinus pinea    Pino domestico 174
19 43479   3       Pinus pinea    Pino domestico 172
20 43480   3       Pinus pinea    Pino domestico 209
21 47504   3       Prunus avium   Ciliegio       59
22 43450   3       Celtis australis      Bagolaro       158
24 43452   3       Celtis australis      Bagolaro       176
25 43453   3       Celtis australis      Bagolaro       171
26 43454   3       Celtis australis      Bagolaro       126
28 43456   3       Celtis australis      Bagolaro       130
29 43457   3       Celtis australis      Bagolaro       127
30 43458   3       Celtis australis      Bagolaro       154
31 43459   3       Celtis australis      Bagolaro       156
32 43460   3       Celtis australis      Bagolaro       90
33 46295   3       Celtis australis      Bagolaro       135
34 46296   3       Celtis australis      Bagolaro       86
35 46297   3       Celtis australis      Bagolaro       90
36 46298   3       Celtis australis      Bagolaro       180
38 46300   3       Celtis australis      Bagolaro       192
39 46301   3       Celtis australis      Bagolaro       169
40 46302   3       Celtis australis      Bagolaro       138
41 46303   3       Celtis australis      Bagolaro       111
42 46304   3       Celtis australis      Bagolaro       95
44 46927   3       Celtis australis      Bagolaro       155
45 46928   3       Celtis australis      Bagolaro       82
49 46932   3       Celtis australis      Bagolaro       196
50 46933   3       Celtis australis      Bagolaro       166
57 47425   3       Robinia pseudoacacia  Robinia 127
58 47426   3       Robinia pseudoacacia  Robinia 245
60 47428   3       Robinia pseudoacacia  Robinia 209
61 47429   3       Robinia pseudoacacia  Robinia 182
62 47430   3       Robinia pseudoacacia  Robinia 212
63 47431   3       Robinia pseudoacacia  Robinia 207
```


Copyright © 2012 ENVIROFI Consortium

9.1.3 NILU AirQUIS Air Quality Data

The NILU AirQuis service provides air quality data in a propriate XML document structure:

XML Format


```xml
<?xml version="1.0" encoding="utf-8"?>
<OnlineDataRoot xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema" Version="v2">
  <Stations>
    <Station Id="7" Owner="Oslo kommune" Name="Alnabru" CoordinateX="10.84664720825650" CoordinateY="59.92752160427910">
      <TimeSeries>
        <TimeSerie Id="41" Component="PM10" Unit="µg/m³" Timestep="3600" DataType="AirQuality">
          <Measurments>
           <Measurment>
              <DateTimeFrom>201101141700</DateTimeFrom>
              <DateTimeTo>201101141800</DateTimeTo>
              <Value>13</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
            <Measurment>
              <DateTimeFrom>201101141600</DateTimeFrom>
              <DateTimeTo>201101141700</DateTimeTo>
              <Value>7</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
            <Measurment>
              <DateTimeFrom>201101141500</DateTimeFrom>
              <DateTimeTo>201101141600</DateTimeTo>
              <Value>10</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
            <Measurment>
              <DateTimeFrom>201101141400</DateTimeFrom>
              <DateTimeTo>201101141500</DateTimeTo>
              <Value>12</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
            <Measurment>
              <DateTimeFrom>201101141300</DateTimeFrom>
              <DateTimeTo>201101141400</DateTimeTo>
              <Value>12</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
            <Measurment>
              <DateTimeFrom>201101141200</DateTimeFrom>
              <DateTimeTo>201101141300</DateTimeTo>
              <Value>13</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
            <Measurment>
              <DateTimeFrom>201101141100</DateTimeFrom>
              <DateTimeTo>201101141200</DateTimeTo>
              <Value>15</Value>
              <QualityControlledData>true</QualityControlledData>
              <Valid>true</Valid>
            </Measurment>
          </Measurments>
        </TimeSerie>
      </TimeSeries>
    </Station>
  </Stations>
</OnlineDataRoot>
```


Copyright © 2012 ENVIROFI Consortium

9.1.4 ERDDAP sea surface temperature data

```
OPeNDAP Dataset Attribute Structure (DAS) format
http://docs.opendap.org/index.php/UserGuideOPeNDAPMessages

Attributes {
 s {
  longitude {
    String _CoordinateAxisType "Lon";
    String axis "X";
    String ioos_category "Location";
    String long_name "Longitude";
    String standard_name "longitude";
    String units "degrees_east";
  }
  latitude {
    String _CoordinateAxisType "Lat";
    String axis "Y";
    String ioos_category "Location";
    String long_name "Latitude";
    String standard_name "latitude";
    String units "degrees_north";
  }
  time {
    String _CoordinateAxisType "Time";
    String axis "T";
    String ioos_category "Time";
    String long_name "Detection Time";
    String standard_name "time";
    String time_origin "01-JAN-1970 00:00:00";
    String units "seconds since 1970-01-01T00:00:00Z";
  }
  station_id {
    String cf_role "timeseries_id";
    String ioos_category "Identifier";
    String long_name "Station ID";
  }
  PeakPeriod {
    String ioos_category "Surface Waves";
    String long_name "sea_surface_wave_period_at_variance_spectral_density_maximum";
    String standard_name "sea_surface_wave_period_at_variance_spectral_density_maximum";
    String units "s";
  }
  PeakDirection {
    String ioos_category "Surface Waves";
    String long_name "sea_surface_wave_to_direction";
    String standard_name "sea_surface_wave_to_direction";
    String units "degrees_true";
  }
  UpcrossPeriod {
    String comment "The zero upcrossing period is defined as the time interval between consecu-
tive occasions on which the surface height passes upward above the mean level. Swell waves are
waves on the ocean surface.";
    String ioos_category "Surface Waves";
    String long_name "Sea Surface Swell Wave Zero upcrossing Period";
    String standard_name "sea_surface_swell_wave_zero_upcrossing_period";
    String units "s";
  }
  SignificantWaveHeight {
    String colorBarMaximum "1500.0";
    String colorBarMinimum "0.0";
    String ioos_category "Surface Waves";
    String long_name "Sea Surface Wave Significant Height";
    String standard_name "sea_surface_wave_significant_height";
    String units "cm";
  }
  SeaTemperature {
    String colorBarMaximum "32.0";
    String colorBarMinimum "0.0";
    String ioos_category "Temperature";
    String long_name "Sea Surface Temperature";
    String standard_name "sea_surface_temperature";
    String units "degree_C";
```

```
  }
 }
  NC_GLOBAL {
    String cdm_data_type "TimeSeries";
    String cdm_timeseries_variables "station_id, longitude, latitude";
    String Conventions "COARDS, CF-1.4, Unidata Dataset Discovery v1.0";
    String geospatial_lat_units "degrees_north";
    String geospatial_lon_units "degrees_east";
    String history "2012-08-21 (source database)
2012-08-21 http://erddap.marine.ie/erddap/tabledap/IWaveBNetwork30Min.das";
    String infoUrl "http://catalog.marine.ie/geonetwork/srv/en/main.home";
    String institution "Marine Institute";
    String license "The Marine Institute complies with the regulations on the Re-Use of Public
Sector Information and we encourage the re-use of the information that we produce.
All of the information featured on our website is the copyright of the Marine Institute unless
otherwise indicated. You may re-use the information on this website free of charge in any for-
mat.
Re-use includes copying, issuing copies to the public, publishing, broadcasting and translating
into other languages. It also covers non-commercial research and study.
Re-use is subject to the following conditions. You must:
Acknowledge the source and our copyright in cases where you supply the information to others;
Reproduce the information accurately;
Not use the information in a misleading way;
Not use the information for the principal purpose of advertising or promoting a particular
product or service.
Not use the information for or in support of illegal, immoral, fraudulent, or dishonest pur-
poses

The Marine Institute is not liable for any loss or liability associated with the re-use of in-
formation and does not certify that the information is up-to-date or error free. The Marine In-
stitute does not authorise any user to have exclusive rights to re-use of its information.";
    String sourceUrl "(source database)";
    String standard_name_vocabulary "CF-18";
    String subsetVariables "station_id, longitude, latitude";
    String summary "Irish Wave Buoys 30 Min";
    String title "Irish Wave Buoys 30 Min";
  }
}

CSV dataset format

longitude, latitude, time, station_id, PeakPeriod, PeakDirection, UpcrossPeriod, Significant-
WaveHeight, SeaTemperature
degrees_east, degrees_north, UTC, , s, degrees_true, s, cm, degree_C
-9.271, 53.227, 2012-08-14T00:05:00Z, Galway Bay Wave Buoy, 2.38, 184.2, 2.685, 52.0, 17.4
-9.271, 53.227, 2012-08-14T00:35:00Z, Galway Bay Wave Buoy, 2.5, 180.0, 2.74, 54.0, 17.45
-9.271, 53.227, 2012-08-14T01:05:00Z, Galway Bay Wave Buoy, 2.44, 174.4, 2.685, 53.0, 17.6
-9.271, 53.227, 2012-08-14T01:35:00Z, Galway Bay Wave Buoy, 2.7, 185.6, 2.759, 57.0, 17.5
-9.271, 53.227, 2012-08-14T02:05:00Z, Galway Bay Wave Buoy, 2.63, 174.4, 2.721, 56.0, 17.5
-9.271, 53.227, 2012-08-14T02:35:00Z, Galway Bay Wave Buoy, 2.78, 187.0, 2.797, 59.0, 17.55
-9.271, 53.227, 2012-08-14T03:05:00Z, Galway Bay Wave Buoy, 3.03, 209.5, 2.837, 62.0, 17.55
-9.271, 53.227, 2012-08-14T03:35:00Z, Galway Bay Wave Buoy, 3.33, 220.8, 2.817, 66.0, 17.5
-9.271, 53.227, 2012-08-14T04:05:00Z, Galway Bay Wave Buoy, 3.03, 218.0, 2.985, 61.0, 17.5
-9.271, 53.227, 2012-08-14T04:35:00Z, Galway Bay Wave Buoy, 3.13, 222.2, 3.008, 56.0, 17.5
-9.271, 53.227, 2012-08-14T05:05:00Z, Galway Bay Wave Buoy, 9.09, 226.4, 3.03, 55.0, 17.5
-9.271, 53.227, 2012-08-14T05:35:00Z, Galway Bay Wave Buoy, 10.0, 223.6, 3.125, 54.0, 17.5
-9.271, 53.227, 2012-08-14T06:05:00Z, Galway Bay Wave Buoy, 3.57, 222.2, 3.125, 55.0, 17.45
-9.271, 53.227, 2012-08-14T06:35:00Z, Galway Bay Wave Buoy, 4.35, 226.4, 3.077, 56.0, 17.45
-9.271, 53.227, 2012-08-14T07:05:00Z, Galway Bay Wave Buoy, 9.09, 230.6, 3.077, 58.0, 17.45
-9.271, 53.227, 2012-08-14T07:35:00Z, Galway Bay Wave Buoy, 9.09, 233.4, 3.125, 55.0, 17.45
-9.271, 53.227, 2012-08-14T08:05:00Z, Galway Bay Wave Buoy, 3.03, 227.8, 2.985, 57.0, 17.45
-9.271, 53.227, 2012-08-14T08:35:00Z, Galway Bay Wave Buoy, 3.23, 222.2, 3.03, 57.0, 17.5
-9.271, 53.227, 2012-08-14T09:05:00Z, Galway Bay Wave Buoy, 9.09, 234.8, 3.077, 56.0, 17.45
```


Copyright © 2012 ENVIROFI Consortium

9.1.5 EUMETSAT sea surface temperature data

```
The EUMETSAT comes as binary data but here are decoded examples in ASCII Grid and CSV format
for sea surface temperature.

CSV format [timestamp, code, correction level, longitude, latitude, sea surface temp]

"2012-01-01 04:00:00","WTMP","no_level",-16.475,50.025,285.49
"2012-01-01 04:00:00","WTMP","no_level",-16.425,50.025,285.96
"2012-01-01 04:00:00","WTMP","no_level",-16.375,50.025,285.81
"2012-01-01 04:00:00","WTMP","no_level",-16.325,50.025,286.07
"2012-01-01 04:00:00","WTMP","no_level",-16.275,50.025,286.08
"2012-01-01 04:00:00","WTMP","no_level",-16.775,50.075,285.48
"2012-01-01 04:00:00","WTMP","no_level",-16.725,50.075,285.01
"2012-01-01 04:00:00","WTMP","no_level",-16.675,50.075,285.7
"2012-01-01 04:00:00","WTMP","no_level",-16.475,50.075,285.94
"2012-01-01 04:00:00","WTMP","no_level",-16.425,50.075,285.98
"2012-01-01 07:00:00","var10_192_2","no_level",-5.925,56.875,12
"2012-01-01 07:00:00","var10_192_2","no_level",-7.475,56.925,-18
"2012-01-01 07:00:00","var10_192_2","no_level",-7.425,56.925,-3
...

ASCII Grid format

ncols 2400
nrows 2400
xllcenter -59.975000
yllcenter -59.975000
cellsize 0.050000
NODATA_VALUE 9.999e20
999900026055400882176.000000
999900026055400882176.000000
999900026055400882176.000000
999900026055400882176.000000
999900026055400882176.000000
286.08
285.49
285.96
285.81
286.07
286.08
285.48
285.01
...
```

9.1.6 AIS shipping reports

```
Example AIS XML shipping alert

<?xml version="1.0" encoding="utf-8" ?>
<ais-binary-message version="1.0">
   <message name="shipdata" aismsgnum="5">
   <description>Class A vessel data report</description>
   <category>F</category>
   <priority>4</priority>
   <operationmode>AU</operationmode>
   <operationmode>AS</operationmode>
   <accessscheme>RATDMA</accessscheme>
   <accessscheme>ITDMA</accessscheme>
   <comstate>N/A</comstate>
   <stationtype>mobile</stationtype>
   <field name="MessageID" numberofbits="6" type="uint">
           <description>AIS message number. Must be 5</description>
           <required>5</required>
   </field>
   <field name="RepeatIndicator" numberofbits="2" type="uint">
           <description>Indicated how many times a message has been repeated</description>
           <unavailable>0</unavailable>
           <lookuptable>
```


Copyright © 2012 ENVIROFI Consortium

```
                  <entry key="0">default</entry>
                  <entry key="3">do not repeat any more</entry>
          </lookuptable>
          <testvalue>1</testvalue>
  </field>
  <field name="UserID" numberofbits="30" type="uint">
          <description>Unique ship identification number (MMSI)</description>
          <testvalue>1193046</testvalue>
  </field>
  <field name="AISversion" numberofbits="2" type="uint">
          <description>Compliant with what edition. 0 is the first edition.</description>
          <unavailable>0</unavailable>
          <testvalue>0</testvalue>
  </field>
  <field name="callsign" numberofbits="6" arraylength="7" type="aisstr6">
          <description>Ship radio call sign</description>
          <unavailable>@@@@@@@</unavailable>
          <testvalue>PIRATE1</testvalue>
  </field>
  <field name="name" numberofbits="6" arraylength="20" type="aisstr6">
          <description>Vessel name</description>
          <unavailable>@@@@@@@@@@@@@@@@@@@@</unavailable>
          <testvalue>BLACK PEARL@@@@@@@@@</testvalue>
  </field>
  ...
```

9.1.7 UBIMET SOS Data Service

```
%obs = (
  'gml:id' => 'org:AUS_BOM:Rundle Island',
  'xmlns:swe' => 'http://www.opengis.net/swe/1.0.1',
  'xmlns:gml' => 'http://www.opengis.net/gml',
  'om:member' => {
    'om:Observation' => {
      'om:procedure' => {
        'xlink:href' => 'org:AUS_BOM:Rundle Island'
      },
      'om:observedProperty' => [
        {
          'xlink:href' => 'urn:ubimet:def:propoerty:AUS_BOM:T_DB'
        },
        {
          'xlink:href' => 'urn:ogc:data:feature'
        },
        {
          'xlink:href' => 'urn:org:def:property:OGC:Time:iso8601'
        }
      ],
      'om:samplingTime' => {
        'gml:TimePeriod' => {
          'gml:end' => {
            'gml:TimeInstant' => {
              'gml:timePosition' => {}
            }
          },
          'gml:begin' => {
            'gml:TimeInstant' => {
              'gml:timePosition' => {}
            }
          }
        }
      },
      'om:result' => {
        'swe:DataArray' => {
          'swe:values' => ',,',
          'swe:elementType' => {
            'swe:DataRecord' => {
              'swe:field' => [
                {
                  'name' => 'T_DB',
```

```
                  'swe:Quantity' => {
                    'swe:uom' => {
                      'code' => 'degree Celsius'
                    },
                    'definition' => 'urn:ubimet:def:propoerty:AUS_BOM:T_DB'
                  }
                },
                {
                  'name' => 'feature',
                  'swe:Text' => {
                    'definition' => 'urn:ogc:data:feature'
                  }
                },
                {
                  'swe:Time' => {
                    'definition' => 'urn:org:def:property:OGC:Time:iso8601'
                  },
                  'name' => 'Time'
                }
              ]
            },
            'name' => 'org:AUS_BOM:Rundle Island_Type'
          },
          'swe:encoding' => {
            'swe:TextBlock' => {
              'tokenSeparator' => ',',
              'blockSeparator' => '@@',
              'decimalSeparator' => '.'
            }
          },
          'swe:elementCount' => {
            'swe:Count' => {
              'swe:value' => '1'
            }
          }
        }
      },
      'om:featureOfInterest' => {
        'gml:FeatureCollection' => {}
      },
      'gml:BoundedBy' => {
        'gml:Envelope' => {
          'srsName' => 'urn:ogc:def:crs:EPSG:6.5:4326',
          'gml:upperCorner' => 'None None',
          'gml:lowerCorner' => 'None None'
        }
      }
    }
  },
  'xmlns:om' => 'http://www.opengis.net/om/1.0',
  'xmlns:xlink' => 'http://www.w3.org/1999/xlink',
  'gml:BoundedBy' => {
    'gml:Envelope' => {
      'srsName' => 'urn:ogc:def:crs:EPSG:6.5:4326',
      'gml:upperCorner' => 'None None',
      'gml:lowerCorner' => 'None None'
    }
  },
  'gml:name' => 'org:AUS_BOM:Rundle Island'
);
%geometry = (
  'SOS=HASH(0x205fcb90)' => {
    'gml:id' => 'org:AUS_BOM:Rundle Island',
    'xmlns:swe' => 'http://www.opengis.net/swe/1.0.1',
    'xmlns:gml' => 'http://www.opengis.net/gml',
    'om:member' => {
      'om:Observation' => {
        'om:procedure' => {
          'xlink:href' => 'org:AUS_BOM:Rundle Island'
        },
        'om:observedProperty' => [
          {
            'xlink:href' => 'urn:ubimet:def:propoerty:AUS_BOM:T_DB'
```


FUTURE INTERNET PPP

Copyright © 2012 ENVIROFI Consortium

```
          },
          {
            'xlink:href' => 'urn:ogc:data:feature'
          },
          {
            'xlink:href' => 'urn:org:def:property:OGC:Time:iso8601'
          }
        ],
        'om:samplingTime' => {
          'gml:TimePeriod' => {
            'gml:end' => {
              'gml:TimeInstant' => {
                'gml:timePosition' => {}
              }
            },
            'gml:begin' => {
              'gml:TimeInstant' => {
                'gml:timePosition' => {}
              }
            }
          }
        },
        'om:result' => {
          'swe:DataArray' => {
            'swe:values' => ',,',
            'swe:elementType' => {
              'swe:DataRecord' => {
                'swe:field' => [
                  {
                    'name' => 'T_DB',
                    'swe:Quantity' => {
                      'swe:uom' => {
                        'code' => 'degree Celsius'
                      },
                      'definition' => 'urn:ubimet:def:propoerty:AUS_BOM:T_DB'
                    }
                  },
                  {
                    'name' => 'feature',
                    'swe:Text' => {
                      'definition' => 'urn:ogc:data:feature'
                    }
                  },
                  {
                    'swe:Time' => {
                      'definition' => 'urn:org:def:property:OGC:Time:iso8601'
                    },
                    'name' => 'Time'
                  }
                ]
              },
              'name' => 'org:AUS_BOM:Rundle Island_Type'
            },
            'swe:encoding' => {
              'swe:TextBlock' => {
                'tokenSeparator' => ',',
                'blockSeparator' => '@@',
                'decimalSeparator' => '.'
              }
            },
            'swe:elementCount' => {
              'swe:Count' => {
                'swe:value' => '1'
              }
            }
          }
        },
        'om:featureOfInterest' => {
          'gml:FeatureCollection' => {}
        },
        'gml:BoundedBy' => {
          'gml:Envelope' => {
            'srsName' => 'urn:ogc:def:crs:EPSG:6.5:4326',
```


Copyright © 2012 ENVIROFI Consortium

```
            'gml:upperCorner' => 'None None',
            'gml:lowerCorner' => 'None None'
          }
        }
      }
    },
    'xmlns:om' => 'http://www.opengis.net/om/1.0',
    'xmlns:xlink' => 'http://www.w3.org/1999/xlink',
    'gml:BoundedBy' => {
      'gml:Envelope' => {
        'srsName' => 'urn:ogc:def:crs:EPSG:6.5:4326',
        'gml:upperCorner' => 'None None',
        'gml:lowerCorner' => 'None None'
      }
    },
    'gml:name' => 'org:AUS_BOM:Rundle Island'
  }
);
```

9.1.8 UBIMET WFS


```
{ "type"      : "FeatureCollection",
  "features" : [
     { "type"           : "Feature",
       "geometry_name" : "position",
       "geometry"      : { "type" : "Point", "coordinates" : [16.2746,48.2687] },
       "properties"    : {
               "VV" : 42,
               "timepoint" : "2012-08-09T09:00:00"
       }
     },
     { "type"           : "Feature",
       "geometry_name" : "position",
       "geometry"      : { "type" : "Point", "coordinates" : [16.2746,48.2687] },
       "properties"    : {
               "VV" : 42,
               "timepoint" : "2012-08-09T09:05:00"
       }
     }
}
```


Copyright © 2012 ENVIROFI Consortium

9.2 Appendix B - Geo-spatial ontologies and information model examples

In the following we present several examples for data and information models for the (Environmental) Observation Web. Our intent is to give an impression of the extent and shape of the state of the art and also point to sources with more complete information.

9.2.1 O&M in INSPIRE

As stated previously, O&M is also used within INSPIRE, here particularly restricting some of the degrees of freedom given by ISO and OGC, but also providing some patters for common usage, for ex-ample for modelling complex phenomena that are measured (see Figure 2 below). An overview of the proposed restrictions and patterns is available from [56].

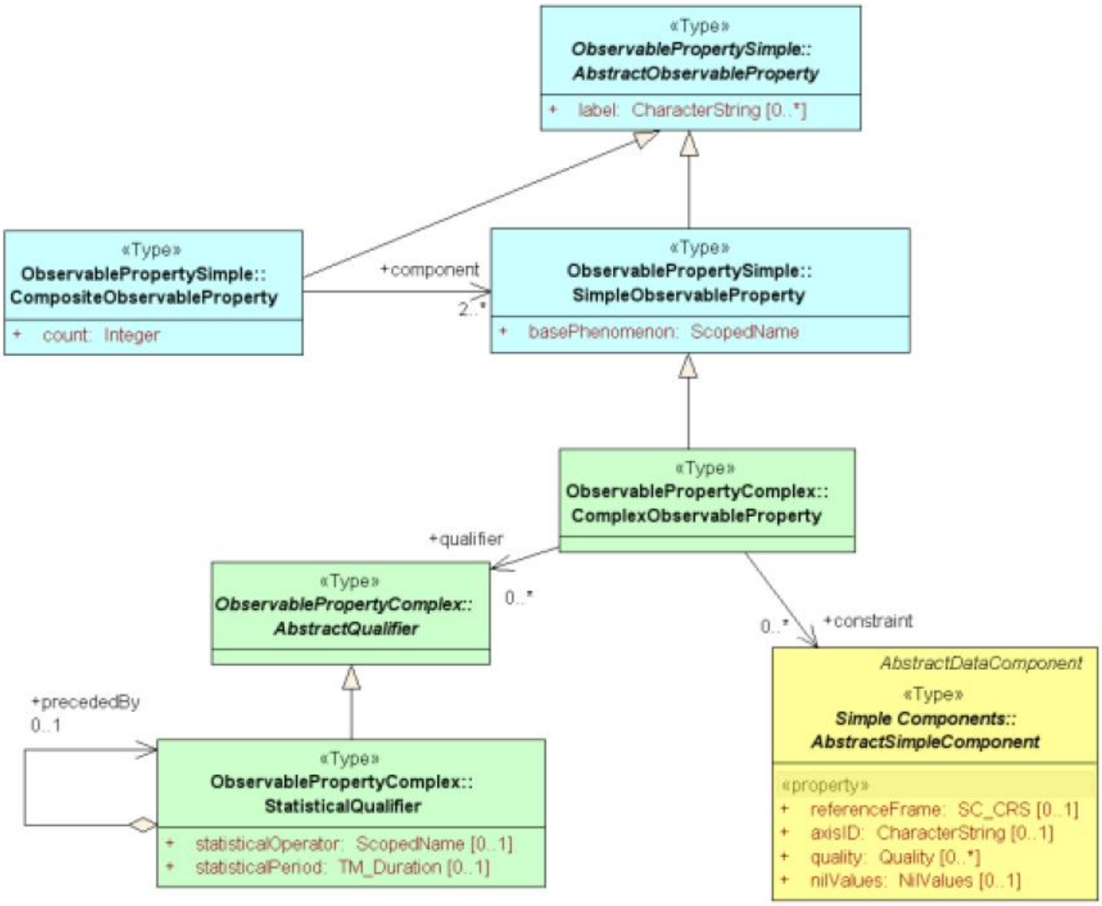

Figure 2. Complex property model for O&M as suggested by INSPIRE [56]

The suggested application of O&M has been considered in the INSPIRE data specifications on:

- Geology
- Oceanographic geographical features
- Atmospheric conditions and Meteorological geographical features
- Environmental monitoring facilities
- Soil

Copyright © 2012 ENVIROFI Consortium

9.2.2 O&M in ENVIROFI WP1 on Biodiversity

Apart from considering the INSPIRE guidelines on the use of O&M, WP1 also ensured that an elaborated information model for species occurrences is facilitated. In this particular case, the TaxMeOn Ontology has been extended to accommodate information about particular organisms (see ENVIROFI deliverable D1.3.2 and also Figure 3 below) and required instances have been created as part of a species knowledge base.

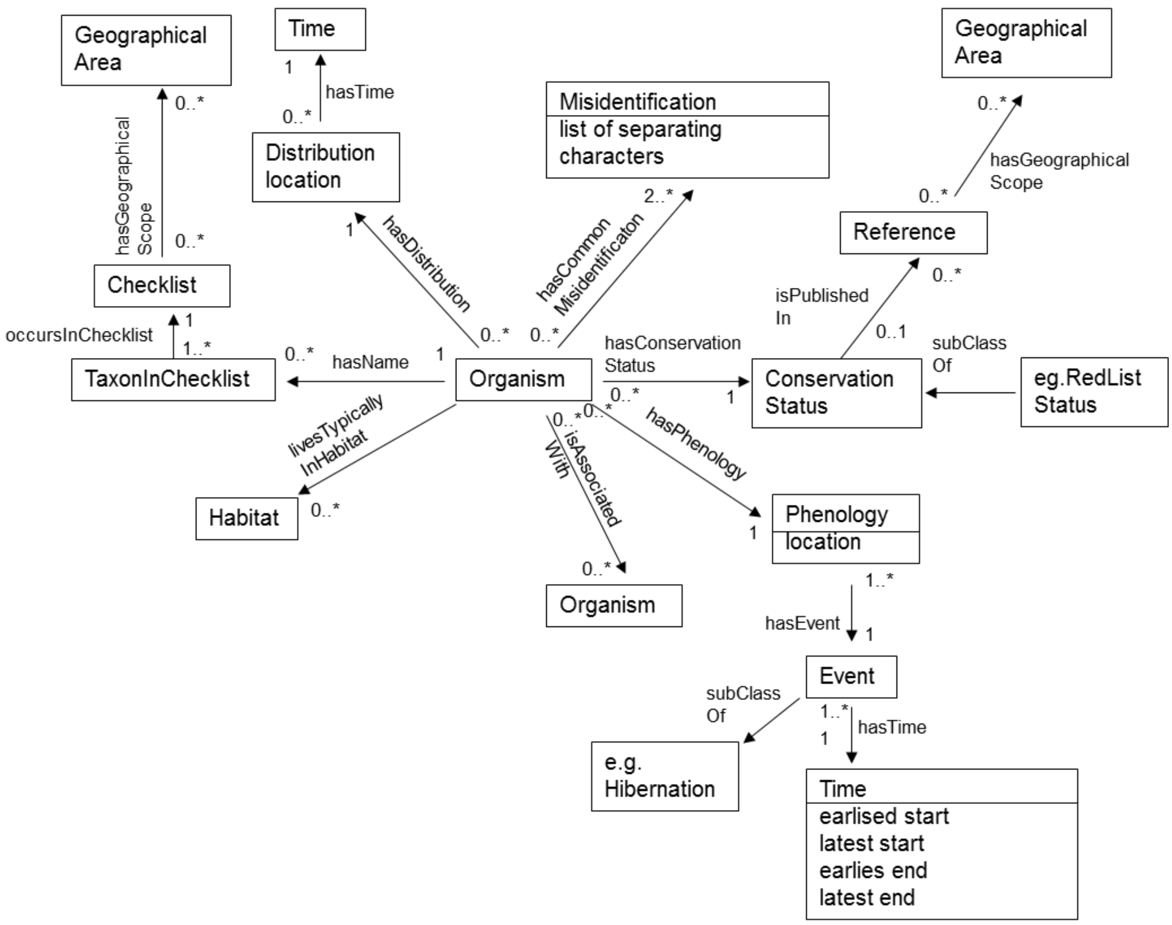

Figure 3. ENVIROFI TaxMeOn Extensions – WP1 (see also ENVIROFI deliverable D1.3.2)

9.2.3 Model for Linked Sensor Data

A linked data model for observation data, following the SSN ontology has been proposed in [63] and extended to aggregated (or fused) observations in [3]. URI Schemas have been suggested for both. The two figures below provide a brief overview. The most important information items include:

- FeatureOfInterest: the entity that comprises observable properties; for example, a 3-dimensional body of air or a sampling point where measurements are taken.
- ObservedProperty: the property that inheres in a feature of interest; for example, the temperature of a 3-dimensional body of air.

- ObservationCollection: a set of observations that is grouped by a distinct criteria; for example, all observations performed by a particular sensor, or all observations of a particular observed property that have been performed within a particular time frame (sampling time).
- Observation: a (social) construct that connects observed properties with sensors, sensing results, and sampling times; for example, the connection between air temperature, a particular temperature sensor, 11.00am (as sampling time) and 23 (as result) degree Celsius.
- SamplingTime: The time instant or interval at which an observation was made; for example, 11.00am in the observation above.
- Result: a symbol representing an observed value; for example, 23 in the observation above.
- Sensor: an entity that performs observations and produces results in form of values; for example, a device that measures air temperature. Humans can also act as sensors.

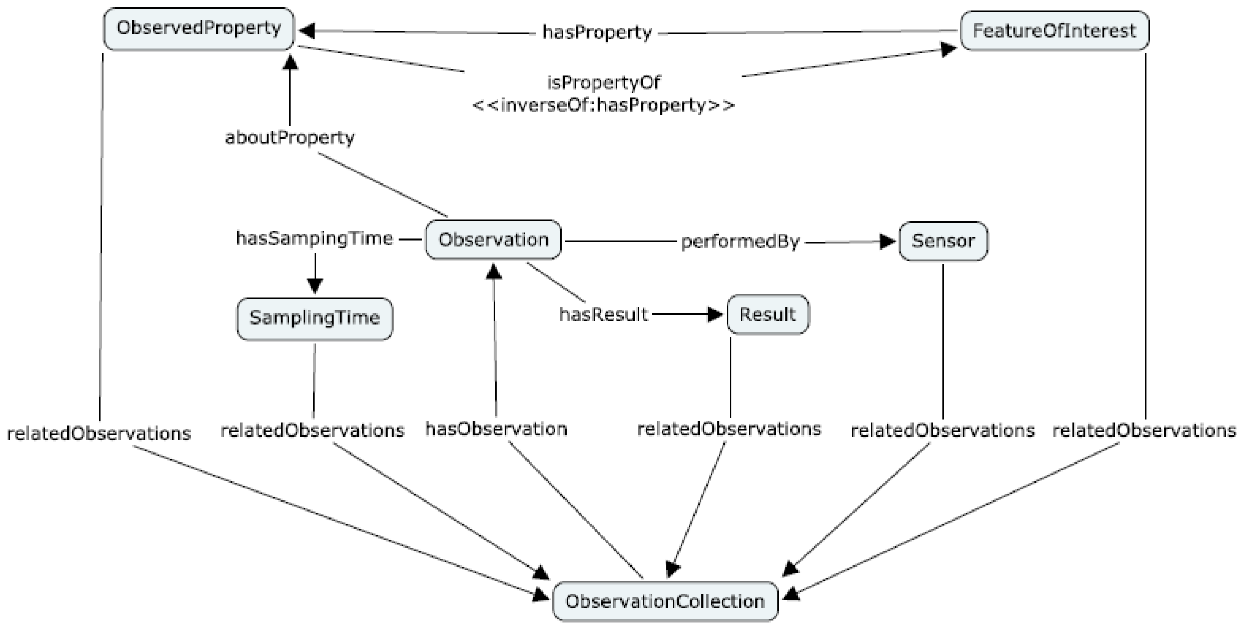

Figure 4. Suggested Linked Sensor Data Model [63]

- isAggregateOf : a relation that allows one observation to be aggregated out of others; e.g., an observation of daily PM10 concentration being an aggregate over hourly measures, or an observation of PM10 in Muenster, Germany being an aggregate over various Point of Interest (POI) measures.
- SensingDevice: a sensor, which is a physical measuring device; e.g., a particular air sampler including a special filter PM10.
- AggregationProcess: a sensor, which implements a concrete aggregation procedure (see below), for example the process that calculates regional PM10 concentrations based on several PM10 concentration observations and additional calibration parameters.
- AggregationProcedure: the specific procedure used for aggregating several observations into one; e.g., calculating the MEAN of 24 hourly observations of PM10 concentration, or a Kriging interpolation method.

Copyright © 2012 ENVIROFI Consortium

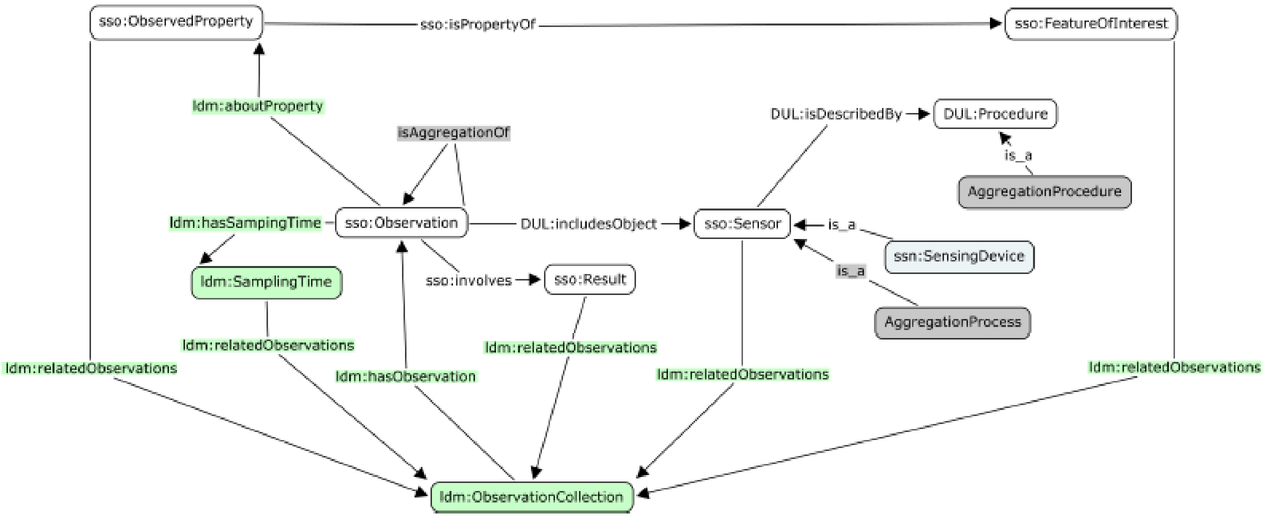

Figure 5. Suggested extension to the Linked Sensor Data model accounting for aggregates [3]

Copyright © 2012 ENVIROFI Consortium