Central Portal as a Data Gateway for EMODnet: Preparations Update

Central Portal Technical Team (VLIZ and EMODnet Secretariat)

Document history

Date	Version	Summary
2020-03-27	0.1	Migration plan- Internal document
2020-12-01	1	Drafting of document

Central Portal as a data gateway for EMODnet

In early 2020 EASME and DG-MARE requested VLIZ to come up with a plan for the migration of the EMODnet thematic lots to the Central Portal. Following continuous discussions and work with the EMODnet secretariat team, EASME and DG-MARE requested a report, to be submitted in November 2020.

The current document follows on from the Central Portal Data Gateway assessment and introduces a more detailed analysis of the lots' infrastructure and presents an overview of the future infrastructure the Central Portal technical team (VLIZ+EMODnet Secretariat) has come up with. As, at this point, the outcome of the phase IV tender is not yet known for all lots, the document deliberately does not include the finer details related with the future Central Portal, especially where it concerns the specific details for each thematic lot and how these will be addressed.

Contents

1   Status Overview

In the previous document (Migration plan submitted to EASME and DG-MARE by VLIZ in early 2020-internal document) an overview of the main features, information and services provided by each lot's portal was collated. It was clear that there was a big discrepancy in how data was handled and made available by each thematic lot.

During the summer 2020, the EMODnet Secretariat organised several meetings where each lot would be able to present an overview of their systems and future developments, as envisaged in phase III contracts, for the benefit of the Central Portal technical team, but also as an opportunity to engage with the different technical coordinators, EASME and DG-MARE. Following each presentation, there were more in-depth follow up meetings with several thematic lots whenever questions arose from those presentations.

Additional to these actions, from September 2020 onwards, the Central Portal technical team started to meet on a regular basis (twice a month) to discuss and progress on the tasks pertaining to the migration process. Even though some discussion was held on possible infrastructure and systems that could be implemented in the future EMODnet Central Portal, the bulk of the work has been directed towards the repatriation to the Europa domain as it constitutes the first deadline in the whole process.

A number of other concurrent actions have also taken place:

- Setup of a GeoHealthCheck instance to monitor the thematic lots' OGC web services uptime and availability (VLIZ)
- Delivery of the EMODnet Indicators Dashboard for the collection of monitoring Indicators information used for all the thematic lots reporting (Trust-IT)
- Setup of the EMODnet JIRA and Confluence instance that replaces the Europa instance (EMODnet Secretariat, VLIZ)
- Setup of a staging/development server so that Trust-IT can implement the changes in the Central portal (Drupal) for the repatriation (VLIZ)
- Setup of an ERDDAP test server (VLIZ)

2   Repatriation

Repatriation is the moving of the EMODnet Central Portal to europa.eu (EC) domain. This is an exercise of changing/applying some DNS settings (in essence). The hosting of the Central Portal website will not move and will remain on VLIZ servers. However, it brings along some obligations in terms of privacy, security, design etc. The repatriation process is a core part of the centralisation process of EMODnet and is dependent on the following tasks:

- **Europa subdomain request:** This is being requested from DG. COMM by DG MARE (in progress)
- **WebGuide/Europa Component Library (ECL):** The design implications to the EMODnet Central Portal are being catered for by Trust-IT. The work is expected to start in December 2020 and should be finished by February 2021.
- **Global Banner/GDPR (cookie consent kit):** These libraries are being integrated by VLIZ, in cooperation with Trust-IT.
- **Europa Analytics:** VLIZ will be implementing Europa Analytics, replacing the current Matomo instance, hosted by VLIZ.

All these tasks have been entered into the EMODnet JIRA and the tickets are being used to track the process. JIRA is the central place to track the repatriation process.

3 INSPIRE compliance

INSPIRE compliance is a contractual obligation for all EMODnet lots and several actions are ongoing and planned to achieve this across all lots. See Table 1below.

Table 1 Ongoing INSPIRE compliance actions

Summary	Status	Brief description
INSPIRE Web Services MetadataUrl and DataUrl fields	Ongoing	Facilitating machine to machine connections is a key element for the development and dissemination of EMODnet (section 1.3 and 1.4.1 of tender specifications). In order to facilitate discovery and usage of OGC Web services (WMS, WFS, WCS), the INSPIRE metadata fields (metadata url - pointing to an xml end-point record - and data url - pointing to predefined download link) should be filled.
Implementation of INSPIRE Compliant CSW service for thematic portals	Ongoing	The Central Portal GeoNetwork harvest metadata from all the lots in different ways. No proper coordination has been achieved so far. Some lots (Physics, Geology, Chemistry) have a catalogue service based on GeoNetwork, some others do not (Human Activities, Bathymetry, Biology, Seabed Habitats). For some lots, the metadata are harvested from their view services and some others provide INSPIRE XML metadata records. All thematic portals need to provide CSW endpoints that can be harvested so that the GeoNetwork can be kept up to date in a more automated way.
Implement INSPIRE performance, capacity and availability normalized testing procedures for view, download and discovery services	Proposal	INSPIRE sets a number of Quality of Service requirements for web services and specifies normalised testing procedures. GeoHealthCheck is an open-source project for monitoring OGC web services uptime and availability. It can be easily expanded and customized with PlugIns. Currently however, no plugins appear to be available to perform the normalized testing procedures as specified by INSPIRE. We propose to expand GeoHealthCheck to implement the INSPIRE normalized testing procedures to monitor the EMODnet web services for INSPIRE compliance. Collaboration with JRC will be sought for the content validation
Make the INSPIRE Validator tool fit-for-purpose for EMODnet	Planned	There is an official EU commission web tool for testing the INSPIRE compliance of web services and metadata. However, the output is difficult to interpret and the tool could be improved. Therefore, we plan to 1. build a more user-friendly user interface that allows anyone to easily view the INSPIRE compliance percentage 2. select the minimum tests for which the EMODnet lots should pass, as not all tests are equally important or relevant 3. make it easier for the lots developers to see what the problem is with a failed test

Alongside these actions, the Central Portal Technical Team (CPTT) has been assessing the best ways to check the compliance for each lot.

The INSPIRE validator (https://inspire.ec.europa.eu/validator/about/), co-developed by Bilbomatica, is the official tool for the EU Commission. Biblomatica are already part of the Secretariat team so it was the logical option to follow up.

An idea was put forward to Bilbomatica to develop a version of this validator that would fit EMODnet's requirements (see above). This point was put forward during the 8th EMODnet Technical Working group meeting and contributions from several participants highlighted the difficulties in implementing full INSPIRE compliance. Further discussions with the EU will be required in order to achieve a priority list of checks that fulfil the contractual obligations but also assist all Technical Coordinators in assessing the compliance for their lots. The list of checks will be defined throughout the migration process and once all the thematic portals have completed this process the list of checks will be used as a reference to the EMODnet INSPIRE compliance.

4   Central Portal Infrastructure design

The EMODnet Secretariat has drafted an initial proposed architecture for creating a central cache for the data and products, as published by each of the thematic portals. This proposal aims to combine existing open-source technologies that are being used in both the data and climate science (ERDAPP) as well as GIS communities (GeoServer & PostGIS).

By offering the EMODnet data and data products consistently through a wide array of community protocols (OGC, GridDap, TableDap), the aim is to attract a greater number of users from diverse backgrounds (academia, industry, policy, citizen science and education) who wish to incorporate EMODnet data into their applications or assess the existing level of knowledge and open data prior to planning for data collection activities. Furthermore, accessing data and data products from across the thematic portals through a single user-preferred protocol, will allow users to seamlessly combine and integrate data and data products from different themes.

This proposal can be achieved through different backend configurations which should be further evaluated by means of testing for feasibility and performance.

One possible configuration (Figure 1) builds on the infrastructure of the existing Thematic Portals web services (OGC, ERDDAP, THREDDS, OPENDAP). Each data and data product published in these web services can be added to an ERDDAP catalogue, which acts as a data broker. ISO metadata for each of these datasets will be harvested by the Central Portal GeoNetwork, already in place, and linked from the ERDDAP catalogue. ERDDAP's pull technology can be used to frequently pull the latest version of the data and data products into the EMODnet Central cache as soon as an update is available or with a maximum delay of two days for real-time data. Data products updates will be slow, due to their intrinsic nature and data updates will occur more often. Note that these updates might be different for the data from the different thematic lots too.

A Geoserver instance will be used to make the data and data products available through OGC services (WMS, WFS and WCS)[1]. The raster data available in the EMODnet Central cache can be published directly in a GeoServer. While vector data in document form can also be published directly in GeoServer, importing those documents in a PostGIS database will improve the performance of WFS and WMS queries. Ideally this import should occur whenever the local cache is refreshed, which

---

[1]   While ERDDAP also supports WMS and WCS protocols, the implementation is not as developed as that in GeoServer, which offers extensions for making these services INSPIRE compliant. ERDDAP does not support WFS.

implies some duplication of storage as document files need to remain available to serve users through ERDDAP.

A second possible backend configuration leverages the common use of PostGIS databases to store vector type data across the thematic portals. These PostGIS databases can be synchronized to the Central Portal cache and can be published directly in the Central Portal GeoServer and ERDDAP (Figure 1). Raster datasets, offered by the thematic portals through simple FTP or the existing ERDDAP/THREDDS/OpenDAP servers, can again be cached using the ERDDAP pull technology and published in GeoServer.

In order to account for vector type data that are not in PostGIS databases (but made available by the thematic portals through OpenDap/TableDAP), ERDDAP can again be used as a broker, followed by an import to a PostGIS Database to publish those layers in GeoServer as well.
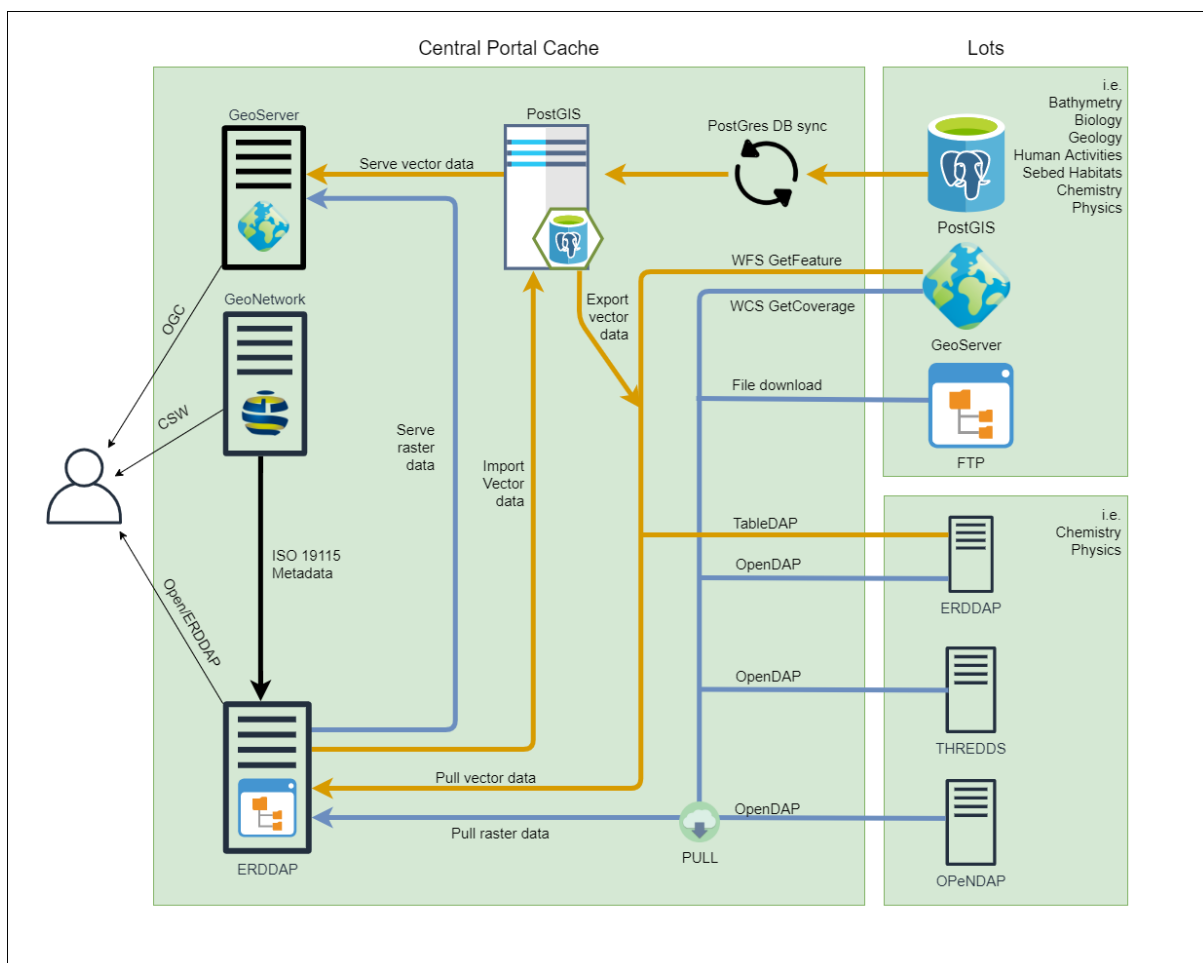

Figure 1: proposed architecture for EMODnet Central Portal

5   Technology Evaluation: ERDDAP server testing

In October 2020, the latest version of the ERDDAP software was used to set up a test server in VLIZ. Initial tests were performed to retrieve data from other systems (i.e. Physics ERDDAP).

Adding data from an existing remote ERDDAP server requires adding the XML that describes the data sets on the remote ERDDAP catalogue to the local ERDDAP servers. The datasets remain on the remote server but are fully accessible from the local ERDDAP servers. Functionality, such as data

searching, format conversion and data sub-setting will all be available via the local server, which passes user requests to the remote server which then executes the commands and returns the results.

Tests were performed with a few NetCDF files to assess the access to remote and local files and were successful through the use of ERDDAP, thus showing that the concept of a Central Portal cache for this type of files can work.

6   Existing Portals Technology overview

The Central Portal Technical team has been recording technologies used by the various portals in the EMODnet community. This process is ongoing. In **Appendix 1** we present the functional technologies that each thematic lot use to share data.

In Table 2 the information technologies used by the portals to deliver their functional services are summarised.

Table 2 An overview of the main information technologies used by the Portals (evolving)

	Unique	AWS	IIS7	PHP	TomCat	Wordpress	PHP	OpenLayers	GlassFish	GEOWebCache	GeoNetwork	MapServer	Geoserver	SQL_Server	Postgis	THREDDS	ERDDAP	Staging Area
Biology	EurOBIS			x									x	x	x			
Physics			x		x					x			x	x	x	x	x	x
Bathymetry	SeadataNet CDI /Sextant												x		x			
Chemistry	SeadataNet CDI/Sextant				x								x			x		
Seabed Habitats		x											x		x			x
Human Activities				x		x	x	x				x	x		x			
Geology	EGDI					x			x		x							

7   Next steps

As the results from the phase IV tenders are announced in the coming months and contracts signed, the Central Portal Technical Team will initiate conversations with each lot to come up with a process and timeline of how the migration work will take place.

As was mentioned in the initial report, the first aim is to guarantee that the metadata, data and data products from each lot are available through the Central Portal. Basic features in the visualisation tools will be also operational at the end of the 6 months of the start of each contract in order for EASME/DG-MARE to feel confidence that EMODnet outputs can be found, queried and downloaded via the Central Portal after six months from the start date of each thematic lot's phase IV contract.

Initially any visualisation tool that is implemented will include basic features that are common to all lots, with the expectation that it will evolve, allowing for features which are specific to a thematic lot to become available too (i.e. the 3D viewing service in the Bathymetry portal).

This work is planned to adopt an AGILE approach and the lots will be granted access to specific areas in the Central Portal development server in case they need to test any technologies before these are made live.

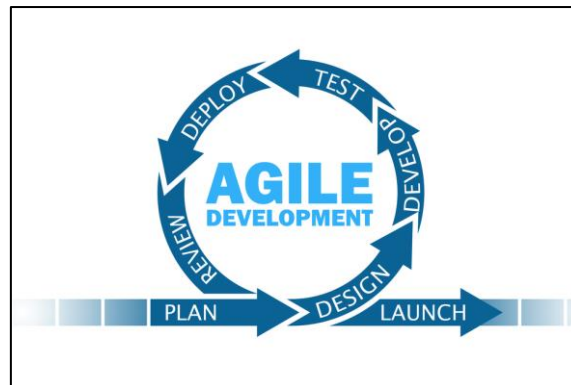Progress will be reported in the Friday meetings between VLIZ, EMODnet Secretariat, DG-MARE and EASME.

Figure 2. AGILE development cycle

EMODnet Bathymetry is the first lot to undergo this process, it is envisaged that the cycle will be refined during the different stages. Once the contract is signed, the Central Portal Technical Team will meet with the Bathymetry technical team and collaboratively devise a plan of works which will be followed up by the design stage where each task is defined and assigned to the relevant people. In the development stage there is an expectation of more in-depth communication between both teams in order to have the output of each task ready for the testing phase. Once this phase is complete, the migration work can be deployed and reviewed. It could be that several tasks happen simultaneously and are at different stages of the AGILE cycle during the first 6 months.

The timeline below exemplifies how the AGILE approach to the thematic lot's migration could be achieved, it is possible that as migration occurs, changes will be required in the timings for each stage of the AGILE approach. One can also envisage that each thematic lot migration will be unique, in that specific stages might differ in length for each lot.
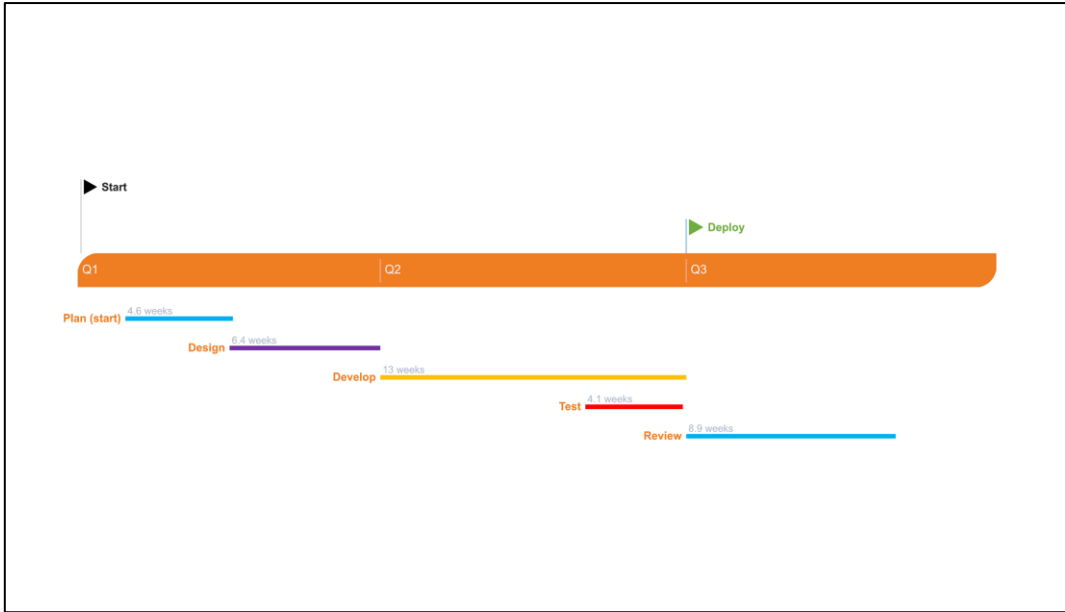
Figure 3 Possible timeline for the development cycle

Appendix 1: An overview of technologies used by the EMODnet community that provide catalogue services.

Metadata search	Data search	Data visualisation	Map visualisation	Data download and sub setting	Data Products	Web services	Other	Usage tracking
Bathymetry								
Catalogue for survey data (Sextant)	Composite Digital Terrain Models (via Sextant catalogue)	Not Applicable	DTM layers 3D visualisation Measure distance Depth profile Add custom WMS Download pre-defined tiles, high resolution area, area of interest	Downloaded from interconnected data centres Survey data not all public domain Layers: average depths, depth contours, quality index for bathy surveys Output formats: Ascii csv, ESRI ascii, NetCDF (CF), GeoTIFF, Fledermaus SD, EMO, .xyz	DTM for European sea regions	OGC compliant (WTMS, WMS, WFS and WCS) REST for depth samples/profiles	Online 3D viewing service Offline 3D viewer	User registration in SeaDataNet

Metadata search	Data search	Data visualisation	Map visualisation	Data download and sub setting	Data Products	Web services	Other	Usage tracking
Biology								
Dataset catalogue (CSW)	Data portal: selection via taxon, dataset, time, region, data precision, parameter	Species occurrences and abundances data, abiotic parameters	Taxa and dataset selection Dataset and layer selection Gridded records of Species distribution	Download toolbox IPT Output formats: DwC, csv,	Atlas of Marine Life (tools, models and spatial maps)	OGC compliant WFS (occurrence and other data) WFS/WMS data products	Data QC tool Online course (contribute data to EMODnet Bio) R Applications	No user registration but data are requested via a form
Chemistry								
Product (metadata) catalogue (Sextant)	Dataset catalogue: selection via region, time, parameters, distributor, country	Dynamic plots for selected groups of parameters (time series or vertical profiles) Select depth/time + animate	Add custom WMS Custom colour map	Not all data public domain Output formats: ascii ODV, raster (PNG), vector (SVG, EPS, PDF), KML, NetCDF, .shp	Spatially interpolated maps of eutrophication parameters in time and depth per sea region, and station time series of contaminants parameters	OGC compliant (WPS, WMS, WFS, CSW) OpenDAP (PyDAP)	Beach and seafloor litter format validators Marine litter manager (Python tool) Live chat	User registration in SeaDataNet (MarineID)

Metadata search	Data search	Data visualisation	Map visualisation	Data download and sub setting	Data Products	Web services	Other	Usage tracking
Geology								
Product catalogue (CSW)	Product catalogue	Not Applicable	Polygon selection Filter functionality (EQ, NEQ, LIKE) Add custom WMS	Downloaded from interconnected data providers or via web services Output formats: .lyr, .shp, containing .sld, .mxd and .qgz files and static maps	Geological layers for European sea regions	OGC compliant (WMS, WFS and CSW)	Not Applicable	No user registration but data are requested via a form
Human Activities								
Product catalogue (ISO19115-2003/Cor.1:2006)	Not Applicable	Aggregated tables (e.g. First Sales of Fish) Elaborate filtering of grouped layers (year, attribute)	Not Applicable	Not all data public domain Output formats: .shp, .gdb, .csv, .mdf, .tif (raster), .xml (metadata)	Human activities (point, line, polygon, raster) for European sea regions	OGC compliant (WMS, WFS, WCS)	No harvesting instance for metadata	No user registration but data are requested via a form (user details mandatory)

Metadata search	Data search	Data visualisation	Map visualisation	Data download and sub setting	Data Products	Web services	Other	Usage tracking
Physics								
Product catalogue (GeoNetwork)	Dataset catalogue	Data plots Trends Data completeness Data dashboard (stats: download, session count, user country, data availability) Route data download for selected platform Time filter Filtering on multiple parameters	Polygon select (coordinates, drawn) Polar projection available Share map link	Core parameters (e.g temperature, pressure, salinity) Some data have links to the providers for download, other can be directly downloadable from portal SensorML (metadata) Output formats: NetCDF, Ascii CSV + GML, GeoJSON, KML, .shp (for route data)	In situ NRT Gridded Climatology	UniqueURL OGC compliant (WMS, WFS) OPeNDAP/Thredds SOAP/REST API ERDDAP Widgets	NRT data Weekly digest	No user registration (except for reprocessed time series)

Metadata search	Data search	Data visualisation	Map visualisation	Data download and sub setting	Data Products	Web services	Other	Usage tracking
Seabed Habitats								
Metadata (layer) search (GeoNetwork)	Not Applicable	Not Applicable	Layers Share map link Zoom to marine region	Data available via web services Output formats: .shp, .gdb, .tif (raster)	EOV habitats OSPAR threatened and/or declining habitats	OGC compliant (WMS, WFS)	Python tools Story maps	No user registration but data are requested via a form)