

General requirements of EMSO

Context of general requirements in EMSO

Complete report on EMSO general requirements available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/41>

Summary of EMSO general requirements

Detailed requirements

EMSO (the European multidisciplinary seafloor & water column observatory) provides research infrastructure for integrating data gathered from a range of ocean observatories and tries to ensure open access to that data to academic researchers. From [1]:

EMSO is a large-scale European Research Infrastructure in the field of environmental sciences. EMSO is based on a European-scale distributed research infrastructure of seafloor observatories with the basic scientific objective of long-term monitoring, mainly in real-time, of environmental processes related to the interaction between the geosphere, biosphere, and hydrosphere, including natural hazards. It is presently composed of several deep-seafloor observatories, which will be deployed on specific sites around European waters, reaching from the Arctic to the Black Sea passing through the Mediterranean Sea, thus forming a widely distributed pan-European infrastructure.

Operation

EMSO is built on largely heterogeneous infrastructure—each institution within the infrastructure operates independently in accordance with its own policies and working practices. Despite this, the basic distributed collection of and access to both data and metadata is already relatively well-realised—the challenge lies with the harmonisation of data formats and protocols, as well as how to provide reliable access to real-time data. EMSO is working towards open access to all datasets. Currently certain datasets are subject to embargos of different lengths, or are simply unavailable for technical reasons.

Technically speaking, the autonomous nature of the various institutions independently deploying observatories and hosting data means a significant heterogeneity of software and technologies adopted for data management, access and processing.

EMSO has to provide services to the community including data provision and the physical access necessary to run specific experiments; it has to be interoperable with similar world-class infrastructures (e.g. OOI in the US, ONC in Canada, DONET in Japan, and IMOS in Australia).

EMSO has facilities for academics to make requests for usage time on observatories, limited to the ocean science community. Technically access to deployed resources is limited to academia rather than industry, as industry may impose data requirements not acceptable to the principles of open access, but this is not a hard restriction.

EMSODEV is a specific (newly initiated) project within EMSO to set up generic 'common instrument packages', making a generic sensor module that can be deployed at various sites and connect to a standard data infrastructure, built based on the principles of the ENVRI reference model.

EMSO is expected to host moorings of FixO³.

EMSO does not currently produce any such events beyond general scientific outreach, but is interested in joint efforts with ENVRI+, especially with regard to use of the ENVRI reference model.

Data and computation

Most observatories contribute data to the MyOcean/Copernicus Marine Environment Monitoring Service. Some data is also contributed to EMECO (the European Marine Ecosystem Observatory). Institutions gather data and links to the data are made available online to researchers. Many observatories store their own data independently of any dedicated data infrastructure; each has its own data management, data access services (typically via FTP). EMSO wishes to promote greater standardisation of data discovery and access.

A goal of EMSO is to harmonise data curation and access, while averting the tendency for individual institutions to revert to idiosyncratic working practices after any particular harmonisation project has finished.

EMSO data may be provided to researchers via different channels. For example seismic data will be channelled via IRIS/EPOS services. Each data domain has different policies, which any unified data infrastructure would have to accommodate those policies.

EMSO is investigating collaborations with data processing infrastructures such as EGI for providing resources for infrastructure-side data processing.

Standards used for data include NetCDF and ODV (Ocean Data View). Use of SWE standards is being encouraged. Metadata generally complies with ISO standards and an extended version of Dublin-Core. It is still necessary however to sometime perform internal conversions of data to be hosted on various services, for example from ISO 19139. EMSO provide open source reformatting software (such as PanFMP) for data not in the desired formats, but this is intended for internal use by individual institutions contributing data.

EMSO has rough data policy written as part of the EMSO ERIC description document, and wants to be able to interoperate with WDS (the World Data System) via long term data archives like PANGAEA. Different data types have different requirements.

It is difficult to normalise data management costs—the independent nature of different institutional nodes leads to distinct differences in the characteristics and emphasis given to data management in general. In terms of volume, the data handled by both research infrastructures is not large (in Big Data terms), though certain data streams (such as provided by subsea camera systems) could be significantly larger if all raw data was retained (which is not currently planned).

EMSO does not collect processing data (from user activities)—however EMSODEV may involve the leasing of resources from external data infrastructures if deemed desirable (this has not been decided as yet).

Based on the ENVRI prototype, OpenSearch is being considered as a means to access distributed data products. OAI is being used internally for data and for data stored in PANGAEA.

Although the general policy is for data access to be open and free, it is currently the case that Copernicus data is password-protected, so a means to bypass this may be required to support open access.

Generally secured/privileged access to data is not a concern of either project due to their umbrella open access policies, however some accounting of data retrieval is a necessary precondition of some institutions contributing their data. Some tracking of data retrieval may also be implemented within EMSODEV.

There is a notable overlap between EMSO and FixO³ data (i.e. some FixO³ data is provided within the EMSO infrastructure).

EMSO requires better mechanisms for ensuring harmonisation of datasets across their distributed networks. Heterogeneous data formats make life difficult for researchers. Improved search is also desirable; currently expert knowledge is required, for example to be able to easily discover data stored in the MyOcean environment.

References

[1] EMSO—European Multidisciplinary Seafloor and water column Observatory, September 2014. <http://www.emso-eu.org/>, accessed 16th September 2015.

Formalities (who & when)

Go-between	Paul Martin and Yin Chen
RI representative	Robert Huber and Andree Behnken
Period of requirements collection	
Status	