EMODnet Jamboree 21/10/2015 Cross Cutting Topic 2 Data Sharing Frictions & Restrictions

Disclaimer: The recommendations, comments and views expressed in this informal report are the views of the participants of the cross-cutting sessions and do not necessarily reflect the views of the EMODnet Secretariat, EMODnet partners nor the European Commission.

Location: De Grote Post Oostende – Meeting Room Telefonie 2 (2nd Floor)

Time: From 18:00 till 19:30/20:00 on Wednesday 21/10/2015

Moderator: Guenter Hoermandinger (EC DG Environment)

One of the basic principles of EMODnet is Open Access to data: to facilitate and provide easy access to harmonised data collected and stored at various locations around Europe, free of restrictions on use. There is also an accelerating trend towards provision of access via machine to machine communication which requires openness and interoperability between data and systems. However, in reality there are considerable challenges in achieving these goals due to tensions between the needs of data collectors and providers (research, commercial, political, security interests) on one hand, and the needs of data users who want fast, easy and free access to the data on the other hand. Also data providers have a need to collect information about who and why the data is being used and this may also restrict easy access. How should EMODnet deal with these different interests, restrictive data policies and complicated log-in procedures? How to strive towards the basic principle of providing open, easy, unrestricted and free access to the data?

Speakers:

- Introduction by Iain Shepherd (EC DG MARE)
- **Thomas Loubrieu** (Ifremer, SeaDataNet, France): Key challenges in reconciling needs and requirements of data providers and data systems versus those of users
- **Conor Delaney** (Joint Research Centre): The power of data brokers and requirements for data openness for machine to machine communication the use case of the Blue Hub facility of the JRC
- **Gisbert Breitbach** (Institute of Coastal Research Helmholtz-Zentrum Geesthacht, Germany): Perspectives on data sharing and licencing issues experiences in Germany

Summary of the discussions

Introduction Iain Shepherd

The European Commission supports open data sharing practices, but recognizes that there are several reasons why people do make data free and available, for example:

- data generators may be hesitant to share data to avoid users making "wrong" conclusions based on their data;
- data generators may want to maintain the integrity of their data and do not want duplicate copies circulating;





- there may be issue over personal data (e.g. in context of fisheries: records of vessel location could be used together/aligned with records of catches to provide personal information on who was fishing where and when);
- data generators may want to be the only one providing the data, and sometimes benefiting from that.

However, a balance can be found and there are few legitimate reasons for hiding or locking data indefinitely.

Thomas Loubrieu (ifremer, SeaDataNet)

Why are data providers reluctant to share data with users?

- Industry / SMEs may lack the human resources to carry out data management.
- Environmental monitoring bodies (Members States) may not want other people to have an opinion on their data.
- Sometimes data is commercially valuable, so the data centre could exploit the data (e.g. for generating products) instead of providing free for users to exploit.
- Sometimes there are specific data policies in place e.g. for certain types of observations in territorial waters.

Possible solutions?

- Sustained funding to ensure sense of security and promote self-confidence for data providers.
- Provide services for data access an data management for free in return for data (e.g. Argo H2020 ?)
- Leverage data against funding: basically if data from a project is not made available then no future funding.

Data use and traceability:

- Open data does NOT mean anonymous usage. Identification of users and tracking download transactions helps to motivate funding organisations.
- Enable promotion of data contributions e.g. DOIs on data sets to measure contribution of data to science and knowledge.
- Standardise data citation by publishers e.g. feedback statistics (spatineo.com, soundcloud.com)
- Cloud paradigm 'free services against information'
- Providers don't want many copies of their data 'out there' unique platform identifiers proper identity.
- Promote a collaborative environment (e.g. Ifremer sensor nanny).

Conor Delaney (Joint Research Centre, European Commission)

A description of the use of marine observation data together with vessel tracking data to predict piracy attacks off the coast of Somalia was provided, and the influence of weather on the incidence of these attacks was highlighted.

The Blue Hub (JRC): an in-house platform to perform research in the fields of maritime surveillance and maritime situational awareness. Blue Hub is a user of marine data. Live feed of global picture – web GIS and Map Server.

Everything is moving cloud-based. This is not because of big data (observation scientists have always dealt with big data), but the reason everything is going to the cloud is because big players such as Apple,





Microsoft etc. are sucking everyone into it. The advantage is that you don't need data storage capacity on a desk-top anymore – huge power available remotely.

ERDDAP (the Environmental Research Division's Data Access Program) was developed to provide a web API for the existing distributed data network.

United States (US) and their data policy:

- The US mainly work with distributed data networks they are not storing data in one place although it may look like it. American data is open by presidential decree.
- The National Ocean and Atmospheric Administration (NOAA): the Data Access Procedural Directive declares that all NOAA environmental data with limited exceptions shall be made discoverable and accessible via the Internet in a timely fashion.
- Machine readable data should NOT require any human intervention.
- The Irish marine institute data license (long and verbose) contrasts with the US license (simple)

ENVIROFI project – Data delivered as web services and composed within the web pages. Can access all info through basic html and javascript live feed from weather buoys and ocean observations.

The Arctic (hot topic): accurate info is needed to deal with emerging challenges and policy response. Norway makes data available via a portal; Canada provides pdfs; US gives data away for free; Russia: not clear.

Gisbert Breitbach (Institute of Coastal Research Helmholtz-Zentrum Geesthacht, Germany)

Described COSYNA – Coastal Observing System for Northern and Arctic Seas

- Data available at various "levels"
 - Level 0 = raw data access offered on private communication
 - Level 3 = geo referenced near real-time data open to all.
- The funders require them to know who is using the data. Therefore they have mandatory login procedure no personal information required
 - o User id
 - o Password
 - o Country
 - o City
 - o Category
 - o Additional voluntary input level
- COSYNA experience shows that the system of collecting information about usage could be reconciled with need to provide users with easy access to the data.

Open floor discussion

- Data-sharing in Europe needs a culture change leading to open-access by default and possibly mandatary as in the US.
- Importance of differentiating between raw data and processed data.
- There is need for use of licenses for downloading data e.g. creative commons.
- EMODnet does not have a standard text for license. Difficulty is that EMODnet is collation of EU data.
- The EC needs to know how data is being used, they need stories: how to achieve without imposing further restrictions and barriers for users?





- We need a more positive message to encourage providers to share their data not just that it is mandatory. Proper crediting/acknowledgement systems are needed.
- EMODnet should not be copying data that is held in HELCOM or OSPAR.
- Duplication of data is problematic: data providers want to provide data once and in the highest possible quality format. Sometimes they are being asked to provided it in an inferior way or in 2-3 different ways. Duplication of data means that updates are not always tied to duplicated versions.
- Sometimes it is necessary to move polygons to make data fit. Data should be made available in its fullest format.
- Is harmonization actually really useful or does it result in a loss in the quality of the data?
- EMODnet should say it will not accept data unless it is free and open!
- Even if providers want to make data available sometimes there are licenses that they are required to get to carry out the survey which then prevent them from passing on the data.
- Need a citation index for researchers depositing data as for publications, linked to H scores this would be motivational for data providers. Could EMODnet find and trace DOIs?

Conclusions and key messages:

- There are many reasons why data providers don't provide data, some legitimate (licenses preventing them, embargo period) some less so.
- The group of data providers which the European Commission (EC) and EMODnet can most influence are research organisations.
- The EC could consider taking a stronger stand making it obligatory that any data collected with public funds MUST be made freely available or no further funding (unless good reason e.g. license).
- EMODnet could refuse to accept data unless it is freely available, however it is sometimes useful for data to be visible, even if it is not free, to prevent duplication of efforts.
- Project proposals should be required to include a statement indicating where and when data will be deposited.
- Data providers (research) could be strongly encouraged to deposit data by developing some sort of a data citation index. Could EMODnet play a role in developing this?
- EMODnet should focus on making available the best quality data and NOT replicating data. Data providers do not want to see multiple copies of their data in circulation.
- EMODnet needs to know how their data is being used via a simple user id log in system.
- There is a need for traceability of data possibly through a unique identifier. Can EMODnet develop this?
- Make it easy for research organisations to deposit data in EMODnet this will be taken on by the Data Ingestion Facility (to be established in 2016)
- We need a positive message to encourage data providers to maintain involved Human activities is establishing Memoranda of Understanding with their data suppliers but there is not one size fits all solution.

