The Industry Marine Data Revolution – Report from the MEDIN Open Meeting

Introduction:

The UK Government has committed, primarily through the Marine Science Co-ordination Committee (MSCC), to opening up access to as much marine data as possible. This commitment is detailed in the UK Marine Science Strategy[1] and has been progressed by the Marine Environmental Data and Information Network (MEDIN), and includes opening up access to industry data.

Recognised as having particular hurdles, the challenge of identifying ways to increase access to marine data collected by industry was taken up by the Productive Seas Evidence Group (PSEG). PSEG commissioned a project to assess existing access to industry-sourced marine environmental data; identify barriers to opening up access to these data; and propose solutions to increase access to non-commercially sensitive data. The project report[2] concludes that, in general, marine industries are positive about sharing marine environmental data and that the infrastructure provided by MEDIN (a network of specialist Data Archive Centres (DACs); a centralised data search portal; and a standardised approach for describing and archiving data) is a successful way to facilitate increased access to industry marine data. However, one of the key barriers identified by the PSEG project was a lack of knowledge within marine industries about how MEDIN and the MEDIN DACs work as a mechanism for sharing data. Indeed one of the report recommendations was:

"Engagement with industry by the public sector needs to be improved, PSEG working with MEDIN can facilitate this by encouraging and informing industry, this could be achieved through a workshop."

MEDIN consequently targeted its annual Open Meeting at the commercial sector, with the aim of sharing knowledge about how MEDIN can facilitate sharing industry marine data as well as providing an open forum to discuss some of the issues that the PSEG report did not cover: how to fund increased access to industry data and what could motivate the commercial sector to share their marine data. This document highlights the key points made by the invited speakers and summarises the conclusions of the discussion sessions.

MEDIN Open Meeting

MEDIN held its annual Open Meeting at the University of Liverpool in London on 9th February 2016, attracting over 60 delegates from across the marine community. The meeting consisted of a series of talks from invited speakers followed by breakout discussion sessions focussed on two themes: funding and motivation.

Invited talks:

Access to Industry Data – What's Possible and Who Benefits, Fiona Miller, ABPmer.

---

[1] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/183310/mscc-strategy.pdf
[2] http://www.gov.scot/Topics/marine/science/MSCC/PSEG/data

Fiona presented results from the recently published report "A Review of Access to Industry Marine Environmental Data", which she was the lead author on.

ABP Marine Environmental Research Ltd and Peter Barham Environment Ltd carried out the review commissioned by PSEG to determine: what data are being collected by a range of marine industries; if and where those data are already being shared; what barriers are seen to opening up access to industry data; and what can be done to remove those barriers. Information was gathered via questionnaires and semi structured telephone interviews targeted at trade bodies and key industry personnel. The study found there are a range of data types collected by the 8 marine industries covered by the study. Two of these industries already actively share marine environmental data (Offshore Wind and Marine Renewables) but there are opportunities to increase data sharing within four of the industries (Aggregates, Cables, Oil and Gas and Ports). The remaining two industries do not generally collect environmental data (Commercial Fisheries and Commercial Shipping). Some common themes emerged across the industries and the main barriers to sharing data were perceived to be commercial confidentiality; data owners wanting control over who can re-use their data; industries working to different standards/formats; lack of knowledge about MEDIN and limited motivation to share data. The solutions proposed in the report are to build on the infrastructure provided by MEDIN and to improve industry awareness of the facilities MEDIN offers.

BGS - Data: donation, management, embargos and delivery, Garry Baker, British Geological Survey (BGS).

In his presentation Garry Baker highlighted the service MEDIN DACs provide to archive data that may be commercially confidential.

BGS is the accredited MEDIN Data Archive Centre for seabed and sub-seabed geology and geophysics and hosts the National Geosciences Data Centre, which holds a complex and diverse range of geoscience data: over 200 TB digital data, ~350km of physical core samples and over 17.5 km paper records and reports. BGS make it easy to add new geoscience data into their databases with a simple online system that captures all relevant information needed to share that data. BGS try to eliminate barriers to sharing data by offering a range of ways of delivering data to their users, including online GIS systems and web services. In addition to traditional geoscience datatypes, it is now possible to download digitised scans, log books and core images from a map based interface. BGS are familiar with managing data from industry and, like all the other MEDIN DACs, have an embargo system in place that makes sure that any sensitive or commercially confidential data they hold has appropriately restricted access. Confidential datasets are reviewed after an agreed time period to determine if they need to remain embargoed.

Safeguarding your biological data with DASSH, Charly Griffiths, The archive for marine species and habitats data (DASSH).

In her presentation, Charly Griffiths demonstrated how MEDIN DACs ensure data owners can maintain control over who uses their data.

The Archive for Marine Species and Habitats Data (DASSH) is the accredited MEDIN Data Archive Centre for marine flora, fauna and habitats. DASSH are able to be flexible in how they receive data and are able to archive a wide range of resources relating to biological data. They offer a variety of

ways for users to acquire data, including Web Accessible Folders (WAF) and online GIS systems. DASSH are used to working with data owners with a range of requirements for how widely data can be shared, acknowledging that some organisations want to retain complete control over who can use their data, whereas others are happy for data to be shared completely openly. In all cases, DASSH data scientists work with the data owners to define the appropriate license to describe who is allowed to access and use each dataset.

Sharing Marine Data – All your concerns alleviated in just 20 minutes, Peter Edmonds, The Crown Estate

In his presentation, Pete Edmonds shared the experiences of the Crown Estate making marine environmental data collected by the Offshore Renewable Energy sector openly available.

The Crown Estate's experience of archiving over 100TB marine industry data from over 2000 surveys, over half of which are publically available, gives them a unique perspective on how commercial organisations respond to sharing their environmental data. As part of their lease agreement, The Crown Estate (TCE) stipulate that marine environmental data collected by the offshore renewable energy sector must be made publically available. Whilst a willingness to share data is a positive step, it can prove inefficient if data are not catalogued or well described. The Crown Estate built a system to house the data they own using the standards and guidance provided by MEDIN, thus ensuring that anyone wanting to reuse their data has all the information they need to do so. Similarly to MEDIN DACs, The Crown Estate embargo and review confidential data where appropriate and work with licenses to clarify who can use the data. To date, over £100m has been spent collecting marine environmental data for offshore wind developments. It costs only a small fraction of that to archive that data for the long term. Three potential funding models to pay for archiving industry data were proposed: pay per archive (current MEDIN model), subscription to MEDIN to cover the costs to archive data and public sector pays in a bid to improve national productivity.

Marine Environmental Data from the Oil and Gas Industry – the SIMORC experience, Colin Grant, CG Metocean Consulting Limited

With over 30 years of working with data in the oil and gas sector, Colin Grant provided an insight into how sharing data has directly benefitted the industry in very practical ways: saving lives and saving money.

Marine environmental data are used by the oil and gas sector when designing new facilities, to generate the statistics required for planning as well as for real-time operations. The Civil Aviation Authority made it mandatory for the oil and gas sector to share the metocean data from their rigs to improve weather forecasting and thus make helicopter flights safer. Imposing regulations to share industry data worked well in this instance because there was a large amount of consultation with industry beforehand. The sector recognises there are significant benefits, including clear business benefits, to the industry to sharing met ocean data. Much of the environmental data owned by the oil and gas sector are collected by contractors who are also contracted to hold the data. However the data are not necessarily actively managed and may no longer be easily accessible if a data format becomes obsolete. BP, Shell and Total now share their data using the System of Industry Metocean Data for the Offshore and Research Communities (SIMORC) to ensure the data are well managed,

quality controlled and archived for the long term. The data owners control whether the data are restricted or non-restricted and pay an annual subscription to maintain the SIMORC service. Many advantages of archiving data using SIMORC were recognised. Two key benefits were saving costs/adding value to earlier investments in data acquisition and preserving corporate memory and data when industry personnel change.

Feedback from breakout sessions on motivation

The discussion groups were in agreement with the report's conclusion that there would need to be a clear driver for the private sector to archive data, and discussed several possible categories of motivation.

Mandatory data sharing: The discussion groups noted that a mandatory requirement to share data had worked well in the case of met-ocean data sharing in the oil and gas sector. However, it was noted that the key to this success was involving the industry in designing the legislation. It was also noted that The Crown Estate's requirement to provide data as a condition of license had worked well, and that data sharing was now business as usual for the sector. There was some support in the discussion sessions for the idea of making data sharing compulsory for some other sectors, as it was perceived that it would level the playing field and was seen as fairer than targeting particular sectors or organisations.

Voluntary data sharing: The attendees also discussed voluntary data sharing as a possible approach for some sectors, and noted that there were existing data sharing 'clubs' where organisations pool data resources. It might be challenging to get voluntary data sharing started as it was perceived that the first organisation/sector to do so could put themselves at a disadvantage.

Data sharing to support education/research: Attendees were aware of some examples of industry willingly sharing data for research purposes, for example aggregate industry data being provided for ecosystem research purposes. However, the groups perceived educational drivers to be less effective.

Efficiency and de-risking operations: It was noted that a standardised format and process for archiving data would create efficiencies where data are currently being extracted from reports. If data in license applications was available in Data Archive Centres and of a perceived quality standard, it might be perceived as more reliable and allow for faster decision making. Data sharing can also minimise risk by providing a more complete picture for operational models and decision making.

Accreditation: the idea of accreditation or a MEDIN data 'kite mark' was discussed. Attendees highlighted the example of the Athena Swan accreditation system in academia, and noted that the scheme had gained real traction over a period of approximately 5 years. Data accreditation could improve public perception.

A more customer driven data service: There was a suggestion that making the MEDIN data service more customer-orientated, in the way that Amazon or another big commercial company would, could help to encourage use of data resources. For example, if a user had a log in and a search history then MEDIN could suggest other data products that might interest them, provide automated

alerts when data they have previously downloaded is updated etc. This type of system might also provide a better understanding of how data are used, which in turn might also help encourage data sharing.

Feedback from breakout sessions on funding

The discussion groups found it difficult to separate the question of how to fund increased access to industry data via MEDIN from the question of what motivates industry to share data. However, several suggestions of ways to fund the long term archival of industry data were made.

Funding model: The funding model that MEDIN has adopted and has had in place for several years has the Data Archive Centres core funded by a parent organisation, typically an organisation with a specific interest in the type of data held at that data centre. This core funding pays for all the systems in place to archive data and serve it out, keep file formats up to date etc. Data owners wishing to archive data may be required to pay a cost per dataset, which depends on the type and complexity of the data to be archived. This covers the cost of getting that specific data into the datacentre and for quality checks etc. Data is then free at the point of re-use. There are differences between the various data centres and indeed some data centres such as UKHO don't currently charge a per dataset cost.

Sponsorship: The MEDIN project is funded by its mainly public sector sponsors. It was proposed that the sponsorship could be broadened to include more industry partners with a one off flat fee similar to the BBC licence fee that could cover data archiving costs. Tiered sponsorship could be considered if different levels of funding were required from different organisations or industries.

Broadening DAC remit: Some of the discussion focused on whether the DACs could have a change of remit that would allow them to generate additional revenue to cover the costs of archiving industry data or provide a specific service to industry. e.g. operating a cloud storage / database system, which industry groups could pay to use operationally, rather than maintaining their own databases and files in multiple locations.

Knowledge exchanges: It was recognised that knowledge exchanges could be used to build connections between MEDIN and commercial sectors, enabling skill and expertise sharing. This could potentially be used in place of a cost per dataset.

Value of data: There was a general consensus that data management is valuable but there was a feeling that some operators may see more value in the long term storage and management of specific types of data. e.g. metocean versus species data. However this is likely to differ between industries.

Concerns: Within the discussion groups there was some concern that it would be difficult to convince the commercial sector to share data via MEDIN if they have to pay to do so. However, the SIMORC experience, where major oil and gas companies pay an annual subscription to share environmental data, suggests this may not be the case, as the overall cost to the industry can be reduced by paying a third party to manage data for the long term.