Approaches to data policy in the marine sector

Marine Environmental Data and Information Network (MEDIN) - Research Project

Short Version of Final Report – December 2010

Version 1.2

Authors:

Neil Pittam, The Crown Estate

Chris Hill, The GeoData Institute, University of Southampton

Steve Saxby, School of Law, University of Southampton

Contents

Prepared by:	Reviewed by:	Approved by:
Neil Pittam Chris Hill Steve Saxby	Jamie Moore	Mike Cowling
Final Version 1.2	Issued on: 10 December 2010	

Executive Summary

The Marine Environmental Data Information Network (MEDIN) commissioned an independent study to investigate and report on approaches to data policy in the marine sector in relation to the wider re-use of marine environmental data. The focus of the study was on public sector data, however, private data holdings are also discussed where appropriate.

A **guidance review** was undertaken by The GeoData Institute and a **data policy audit** was undertaken by The Crown Estate. The data policy audit included interviewing 21 public bodies, 6 private companies and 1 charity. The interviews were undertaken between April and June 2010 and the report captures the status of data policies for this time period. These work packages are reported in this document which is a summary version (29 pages inclusive) of the long version of the final report (110 pages inclusive)(MEDIN, 2010).

In the last 2 years, policies regarding the UK's marine and terrestrial environmental data have seen a relaxation of public sector licence terms in some circumstances and there are examples of improved data sharing and re-use, however, some examples to the contrary were found.   There is scope for improving this situation by standardisation; reducing the variety of licences by adopting common licence terms, providing licence templates to simplify data sharing and re-use, and providing contract templates to achieve standardised data generation. It was also found that with some simple measures public bodies could do more to inform data users as to what is available for no-charge with no restrictions on re-use and what is with charge and restrictions. A lack of clarity of when this applied to certain data was found to confuse and sometimes frustrate those trying to obtain these data when perceived blockages to data were not fully explained.

Developments in the marine environmental data sector including those resultant from the EU's Infrastructure for Spatial Information in Europe (INSPIRE) and Public Sector Information (PSI) Directives, The UK Location Programme (UKLP) activities, MEDIN initiatives and centralised government data sharing polices and initiatives such as data.gov.uk. These initiatives are all driving the improvement of the *status quo*, however, clearer licensing and guidance, co-ordinated data activities and increased funding are required to achieve the level of improvement that is generally considered to be required. For example, a number of public sector data portals for sharing and presenting data for re-use either exist or are under development, and the coordination and potential consolidation of these initiatives should be considered.

The main differences in approach regarding data policy and licensing are between public bodies that undertake commercial activities (e.g. the Met Office, UKHO, Cefas and BGS) and those that do not (e.g. JNCC and Natural England) which results in differences in policy regarding how each organisation works with and regards its data.

Main conclusions:

- Interoperability of data and metadata to be improved
- Common standards, policies and agreements are required
- Sharing and re-use is improving
- General clarification is required (e.g. list Public Task datasets and products; Information Fair Trader Scheme (IFTS); UKLP and INSPIRE)
- Many portals exist; co-ordination and rationalisation would be beneficial

- Some practices (costs and licence terms) should be reviewed
- Funding is required for portals and for some organisations to bring their metadata up to the INSPIRE standard

Key recommendations for the following topics are:

- **Access policies**
  - Put a pan-government marine data plan in place to lever the importance of marine data management
  - Organisations to declare what data they hold is PSI (Public Task) on an asset list
  - Make it easier to locate the correct contacts for obtaining data
  - The provision of no charge, freely available, quality assured national datasets would be of great value to the public and private sector alike
- **Operational approaches**
  - Making existing datasets INSPIRE compliant will require significant resources and effort which is a big issue for some organisations. Support, guidance and tools for expediting this process would be valued
  - Data researchers would prefer as few portals as possible to obtain data in a format they can re-use
  - Make it a contractual requirement for contractors to post data to MEDIN Data Archive Centres (DACs) or similar
- **Best practice initiatives**
  - That a single common metadata standard (INSPIRE compliant) is used across the sector as far as is practicable. Changing metadata standards once adopted is extremely costly and time consuming.
- **Pricing and licensing**
  - Pan-government terms to avoid inappropriately high charges for licensing their data to other government organisations

Glossary

It is important to define the meaning of key terms used throughout this report to provide clarity of meaning. In particular it is important to clarify the meaning of the terms of data "re-use" and "sharing" in the context of this study.

Charge – A cost for the data / the associated licence above nominal handling charges. See also 'Data licensing with a handling charge' below.

Click-use – 'Click-use' licences are for re-using public sector information. Click-use is the term used to describe OPSI online licences for the re-use of Crown copyright information and Parliamentary copyright information. There are two Click-use licences: the Public Sector Information (PSI) Licence (formerly known as the Core Licence) and the Parliamentary Licence (this covers Parliamentary copyright information). There is no charge for the PSI Licence or the Parliamentary Licence. Some Crown copyright material is covered by waiver conditions.

Crown copyright – Copyright material which is produced by employees of the Crown in the course of their duties. Therefore, most material originated by ministers and civil servants is protected by Crown copyright. The Director of OPSI in her role as Queen's Printer has been appointed by Her Majesty the Queen to manage all copyrights owned by the Crown on Her Majesty's behalf. OPSI's Information Policy team licenses on the Queen's Printers behalf.

Data licensing with a handling charge – data requests from public sector bodies which do incur a nominal delivery charge due to the complexity and volume of the request.

Data sharing – Sharing of data between organisations for no charge, with no re-use rights.

Data licensing with no charge – downloading data automatically from a portal or a small manually delivered data requests from public sector bodies, with no charge for the data, the associated licence or delivery.

Derived products – (either using the raw data or sources that are themselves derived from the raw data) are a particularly controversial area. Joint Intellectual Property Rights (IPR) and / or ownership are generally recognised as irrelevant as the important issue is the "rights" given to the licensee to use that data. A particular area of difficulty for licensees to understand is that of "copy derived" and "non copy derived" data.

- Copy derived means that the derived data set includes a copy of the original information as a whole or any substantial part of it (as defined by Copyright legislation and case law) or that the derived data set can be reverse engineered to create a copy of the original information or any substantial part of it.

- Non-copy derived means that the derived data set does not include a copy of the original information as a whole or any substantial part of it (as defined by Copyright legislation and case law) or that the derived data set cannot be reverse engineered to create a copy of the original information.

Metadata – A geospatial metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource. It represents the who, what, when, where, why and how of the resource. Geospatial metadata are used to document geographic digital resources such as Geographic Information System (GIS) files, geospatial databases, and earth imagery.

Public Task – The Public Task of a public sector body are the public datasets which it must provide to fulfil its obligation to the Public. These are typically raw data which have had minimal further processing. This is opposed to the commercial business activities of Trading Funds.

Raw data - Basic data and information, raw data, public sector data.

Re-use – the use of data by persons or legal entities of documents held by public sector bodies for commercial or non-commercial purposes other than the initial purpose related to the public task for which the documents were produced. The exchange of documents between public sector bodies purely in pursuit of their public tasks does not constitute such re-use.

Value added data – Raw data (see above) which has had value added to it by creating bespoke products. It therefore contains Intellectual Property (IP).Further definitions and background information are presented in Annex 1 which is taken from Annex A of the study Announcement of Opportunity.

An overview of the 'actors' involved in UK public sector data is presented below as created by the UK Location Programme to explore some of the business requirements of the Discovery Metadata Service (DMS), UK Location Programme (2010).

Actor	Definition
User	Someone who uses data.
Data Provider	An organisation that has data that they wish to publish and hold the rights to do so.
Publishing Agent	Service provider for Data Providers who do not wish to publish their data themselves. A Publishing Agent may have many Data Provider clients and may provide a range of publishing services, including data hosting, metadata management and web services. A Data Provider may act as a Publishing Agent for other Data Providers.
Data Publisher	Either a Data Provider (publishing directly) or Publishing Agent.
Service Provider	The provider of a web service for a given dataset. Typically the Data Publisher, but may be a third party, e.g. in the case of transformation services.
Data User	A user of the published metadata for the discovery and evaluation of datasets and associated services; and subsequent data services.
Discovery Service Client	A user of the UKLII Metadata Catalogue, who wishes to build it into an application, e.g. an information portal focused around particular datasets.
Coordination Unit Officer	A member of the Location Information Coordination Unit. Roles include assurance and management reporting.

Table 1. Outline use cases, UK Location Programme (2010).

1. Introduction

Data and information are the lifeblood of the knowledge economy, yet within the UK information is highly fragmented with many datasets and sources. There is duplication of information collection in some cases and harmonisation between datasets at different scales is often lacking. Co-ordination between different levels of government and public authorities is generally poor resulting in incompatible information and information systems. Much information is not re-useable and there are various policy, institutional and commercial restrictions on access to data. What data there is can often be difficult to identify, access, use and re-use.

This independent study investigates the approaches to data policy in the marine sector in relation to the wider usage of marine environmental data with the ultimate aim of long-term sustainability of marine datasets. The focus of the study was specifically regarding data policy and licensing issues related to public sector data, however, private data holdings are also discussed where appropriate. The two work packages undertaken were:

- Summary of the status of legislative and regulatory guidance relevant to data policy
- Audit based on interviews was conducted to review:
    - Best practice
    - Compliance
    - Pricing and licensing
    - The user perspective

The objective was to conclude on the guidance review and audit and to make recommendations on how the *status quo* could be improved. Useful recommendations came out of the interviews that do not directly relate to the remit of this study (e.g. relating to metadata and standards), and these are provided in Annex 6 in the full report.

The audit interviews were undertaken between April and June 2010 and the report captures the status of data policies for this time period. The study was completed with the political backdrop of the 2010 General Election which took place approximately half-way through the study (6 May 2010) and saw a Labour government succeeded by a Conservative-Liberal Democrat coalition. This change of government naturally changes the UK's data policy and it is important to note that the full impact of this is unclear at the time of writing. Economically the UK was 16 months into a recession and the new government had already initiated the debate on broad reaching cuts across the public sector. These cuts had the potential to impact upon the majority of those interviewed, therefore, potential cost savings to be made from efficient data policies were extremely pertinent and timely.

This is a summary version (29 pages) of the final report (110 pages, MEDIN, 2010).

2. Scope and purpose

Fit for purpose data and information, which is the most current available, is essential to underpinning evidence based decision making and enabling Government to deliver on its policy objectives for the UK's marine environment and coastal zone. Ensuring the most productive use of

existing and new data and information is a key aim of the UK Marine Monitoring Assessment Strategy (UKMMAS) and the new Marine Science Strategy.

Marine data and information exists in many forms and is acquired, managed, manipulated, and used by a wide range of public and private sector organisations. The manner by which these organisations are funded varies significantly as do the commercial and other interests that these organisations place in the use and re-use of data.

The purpose of this study was to work within the current mixed funding model to generate an independent and balanced picture of the situation with regard to approaches within the Marine Sector to charging for access to data and licensing / charging for re-use. The intention was to identify any areas of the current legislation or guidance that are unclear and provide a balanced view of the issues to allow Government to take / make the necessary judgements and steps to improve access to data overall.

Consideration of the long term economics (particularly in relation to levels of revenue to Government through tax) are outside of scope but the review documents the costs of the current situation as far as they can be determined.

The work packages undertaken were:

1. Guidance Review

   - Independent review the current guidance underpinning both charging and the re-use of public sector information and data with specific focus on marine data.
   - Liaised with OPSI to ensure there is a clear understanding of work being undertaken and OPSI's view on how the current framework should be operating.
   - Liaised with ERFF and UKLP to ensure this work complements and feeds into activities planned by them.
   - Liaised with members of APPSI, as necessary, to advise of the scope of the review.

2. Data Policy Audit

   - Through a series of interviews a detailed audit of a cross section of the MEDIN community was undertaken, including the four current DACs, a section of public sector data users / providers and a representative sample of the private to review:

     (a) Best practice
     (b) Compliance
     (c) Pricing and licensing
     (d) The user perspective

3. Marine data policy guidance review

3.1. Overview

This review covers national and EU legislation related to PSI and the guidance relating to charging and re-use of public sector information from literature and discussions with key UK organisations (OPSI, UKLP, ERFF, and APPSI). It has focused as far as possible on marine data and the charging and re-use of public sector data in this domain. The scope does not allow for the detailed section by section analysis of current law and guidance; partly mediated by the recent reviews of this within the marine sector (MRAG et. al., 2009). Thus the scope of the review in the full report is offered as a framework analysis and literature search of the relevant policies, legislation and guidance, with an overview of trends.

The scope of this review has been from the perspective of MEDIN as a key coordinating body nationally in promoting discovery and access to marine data that seeks to discuss the potential role for MEDIN in clarifying guidance to broadly assist users. MEDIN has recognised the lack of coherence on data exchange, charging and rights management. First we observe that MEDIN is an organisation or rather a network of sponsoring organisations that have come together to pursue the common cause of promoting the sharing of, and improved access to, a range of marine data through a central discovery metadata portal (DMP). MEDIN now wishes to attack some of the ongoing problems faced by data users, which involves in particular trying to make marine datasets more identifiable, accessible and useful as an online resource. Implicit within this must be the desire to supply the right information to the right people at the right time, which is a cornerstone of Government objectives with regard to public sector information (PSI).

The contents of Section 3.1 can be seen in the long version of the final report (MEDIN, 2010).

3.2. UK PSI distribution and re-use policies

Section 3.2 of the final report presents the UK PSI distribution and re-use policies.

4. Data policy audit

4.1. Methodology

The data policy audit was conducted with 28 organisations (27 of those face-to-face) and those interviewed were primarily 'data policy' and 'operational' employees. The organisations interviewed are listed in the acknowledgements.

4.2. Differences in approach to policy

Within the UK public sector marine data fraternity a variety of business models exist. The main differences in approach are between public bodies that undertake commercial activities (e.g. the Met Office, UKHO, Cefas and BGS) and those that do not (e.g. JNCC and Natural England).

The classifications referred to in this report are as follows:

Business model type	Organisation name
Ministerial Departments	Defra, Marine Scotland (both Crown bodies) and NIEA (Department within Department of the Environment (DOE))
Commercial Public Body	The Crown Estate (Crown body)
Next-Step Executive Agencies	CEFAS, TNA (OPSI)* and MCA; (all Crown bodies)
Executive Agency Trading Funds	UKHO and Met Office (both Crown bodies and MoD trading funds)
Non-departmental Public Bodies (NDPB's)	EA, MMO, SNH, English Heritage, CCW, JNCC, Natural England, AFBI and APPSI
Research Institute (Research Council Body)	BGS and BODC

* Merged in 2006.

Table 2. Organisation Types and Classification

Non-public sector organisations who also supply public sector data are:

Business model type	Organisation name
Charities (independent non-profit making organisations)	MBA (DASSH project)

Table 3. Non-public sector organisations who also supply public sector data

In summary the relevant data policies for these organisation types were:

Business model type	Summary data policies
Ministerial Departments	(Defra, Marine Scotland, NIEA) Public Task activities only, noting that Defra do not supply data directly but fund data related activities. All data is public domain (subject to exceptions and exemptions) with no-charge. Licence terms are 'Crown copyright' and 'Click-use'.
Commercial Public Body	(The Crown Estate) Licence terms are 'Crown copyright'.
Next-Step Executive Agencies	(Cefas, TNA (OPSI), MCA) Public Task activities only. All data is public domain (subject to exceptions and exemptions) with no-charge. Typical licence terms are 'Crown copyright' and 'Click-use', with bespoke variations. Cefas also undertake

Business model type	Summary data policies
	commercial activities.
Executive Agency Trading Funds	(UKHO, Met Office) Public Task and commercial activities. Required to generate income. Public Task data is public domain (subject to exceptions and exemptions) with no-charge with typical licence terms such as 'Crown copyright' and 'Click-use'. Commercial activities are subject to bespoke licences and related charges at market rates.
Non-departmental Public Bodies (NDPB's)	(EA, MMO, SNH, English Heritage, CCW, JNCC, Natural England, APPSI, AFBI) Public Task activities only, fully funded, not required to generate income. All data is public domain (subject to exceptions and exemptions) with no-charge. There are some exceptions (e.g. within EA, AFBI) that are referred to in the full report. Typical licence terms are 'Crown copyright' and 'Click-use', with bespoke variations.
Research Institute (Research Council Body)	(BGS, BODC) Typical licence terms are 'Click-use' and NERC data policy, with bespoke variations for commercial licences. Some public domain data.
Charities (independent non-profit making organisations)	(MBA, DASSH project) Data holdings are public domain with no-charge and private with charges. Typical licence terms are 'Crown copyright' and 'Click-use' for public data with bespoke variations for private data holdings.

Table 4. Summary of the relevant data policies for organisation types

When referring to data that organisations share and re-use it is important to make the distinction between raw and higher level, value added products (see glossary). Raw data is the basic dataset which has been processed to be clean, calibrated and standardised. Raw data is the format that public sector data typically takes and forms the basis of any analysis or reprocessing. Value-added data has had value added to it by creating bespoke products from the raw data, which can result in it containing Intellectual Property (IP) which can require a specific licence with associated charges for re-use, depending on the organisation's policy.

Departments, Next-Step Executive Agencies, Non-departmental Public Bodies and Research Institutes

For all but the Executive Agency Trading Funds and Charities (for a small percentage of data holdings which are private or commercially sensitive), the policy on data use, sharing and re-use is one of no-charge and standard licensing, with the general ethos of making these data as widely available and easy to access as possible.

Executive Agency Trading Funds

Executive Agency Trading Funds undertake Public Task and commercial activities and the challenge for them is to clearly differentiate between the two and not overlap them.

Charities

Charities are typically funded to complete their data sharing and re-use activities. Public domain data is provided for no charge and made as widely available as possible.

Private companies

Private companies have no obligations in terms of providing data to public sector bodies but data sharing does occur. Public sector data requirements are followed as per contractual requirements.

Table 5 summarises the general policies for the various types of organisation when acting as a data provider.

Organisation Type	Public Task	Commercial	Standard public data licence with no charge or minimal data handling charge	Commercial licences with market rates
Ministerial and Non-Ministerial Departments	**Yes**	No	**Yes**	No
Next-Step Executive Agencies	**Yes**	No*	**Yes**	No
Executive Agency Trading Funds	**Yes**	**Yes**	**Yes**	**Yes**
Non-departmental Public Bodies (NDPB's)	**Yes**	No	**Yes**	No**
Research Institute	No	No	**Yes**	**Yes**
Charities	No	Mainly no	**Yes**	Mainly no
Private companies	No	**Yes**	No	Yes

* Cefas also undertake commercial activities.

** Some exceptions (e.g. within EA) that are referred to in the long version of the report (MEDIN, 2010).

Table 5. Summary of the general obligations and policies for the various classifications of organisation.

Government data policy initiatives

The previous administration ran a number of initiatives and studies to improve data sharing and re-use:

- The Power of information, Mayo and Steinberg (2007)
- Making Public Data Public (takes forward the Power of Information Taskforce's report)
- Putting the Frontline First: Smarter Government, HM Treasury (2009)
- Cross Cutting Review of the Knowledge Economy' report, HM Treasury (2000)

A key result of these has been Ordnance Survey (OS) releasing certain data for free with the public having more access to OS maps from 2010 as part of a Government drive to open up data to improve transparency. Such data relates to electoral and local authority boundaries, postcode areas and mid-scale mapping information. Similar initiatives have been observed from other public bodies

which have been followed by other organisations such as BGS and UKHO and this trend is anticipated to continue.

4.3. Description of data sources and access policies

4.3.1. Data sources

Table 6 in the full report (MEDIN, 2010) presents the public marine environmental data sources, data sources considered, datasets available, access policies, pricing and compliance with standards for those organisations audited.

Access policies and how they are applied

In addition to the access policies outlined in Table 6 in the full report, this section contains more detail on operational responses related to data policies. Table 7 in the full report summarises how regulations processes regarding data access are being applied at the operational level.

4.3.2. OPSI IFTS members

The OPSI Information Fair Trader Scheme (IFTS) sets and assesses standards for public sector bodies that trade in data. It requires them to encourage the re-use of information and reach a standard of fairness and transparency. The following interviewed organisations are members of OPSI IFTS:

- BGS
- EA
- Met Office
- TNA (OPSI)
- UKHO

Members are required to carry out re-use activities following key principles:

- Maximisation
- Fairness
- Transparency
- Simplicity
- Innovation
- Challenge

More details on IFTS are provided in section 4.4 in the full report. 'Understanding of best practice within the sector'. A full list of members can be seen on the National Archives website.

4.4. Understanding of best practice within the sector

Interviewees were questioned directly regarding their organisation's knowledge and experience of the following:

- INSPIRE
- UKLP

- MEDIN
- FOI / FOISA (FOI Scotland Act)
- EIR / EIRS (EIR Scotland)
- OPSI IFTS

Interviewees were also prompted to flag other initiatives which were relevant to their data practices. Figure 1 presents a summary of the responses and specific comments on best practice initiatives.
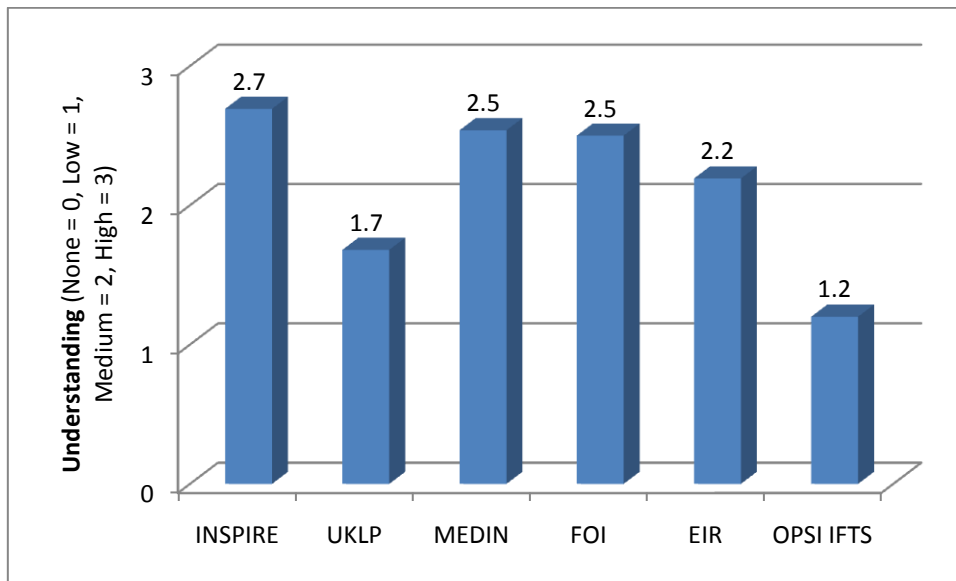
Figure 1. A summary of the understanding of best practice within the sector

Annex 5 of the final report contains supplementary information on organisation's understanding of best practice within the sector.

4.5. Pricing and licensing of data

The principal licence types are:

- Click-use licence (OPSI)
- Click-use licence (data.gov.uk)
- Standard in-house licence for free PSI data
- Standard in-house licence for commercial data
- Bespoke in-house licence(s) for commercial data

A summary of pricing and licensing is presented in Table 9 of the final report.

Specific comments on pricing and licensing from interviewees and an overview of TNA's (OPSI) influential role in these matter is provided in Annex 8 in the final report.

4.6. Markets

In the final report this section details existing data market conditions, changing market needs, volumes and opportunities. This is presented by interviewee where information was available. Market volumes and opportunities are presented in the sub-sections 4.6 in the final report.

5. Initiatives to improve the sharing and re-use of marine environmental data

A number of public sector data sharing portal initiatives exist. To avoid duplication of efforts and inefficient government funding it is important that these activities are coordinated and that each organisation is clear about its remit.

A key benefit of using a data portal is that it allows for data to flow into the archive and be readily accessible. A secure database has benefits over an uncontrolled spreadsheet for example, allowing for the creation of quality assured documents ready for putting into reports. The community can also state methodologies and run the same queries and replicate results. Also, staff turnover can be an issue in terms of the continuity of service which can happen quite frequently and be frustrating to work with. Portals overcome this issue.

MEDIN

MEDIN is promoting the establishment of a network of MEDIN accredited DACs as the recommended places for the archiving of marine data sets to ensure secure long-term storage of data according to a set of best practice principles. These DACs are required to produce and publish metadata records on their data holdings.

The present DACs are:

- Bathymetry – UKHO
- Biodiversity – MBA (DASSH project)
- Geology - BGS
- Oceanographic (physical, chemical, biological and geotechnical) – BODC

and they are all in different states of development and completeness. All are public bodies apart from MBA (DASSH project) which is a charity.

In order to provide assurance to the data provider:

a)   A DAC is required to demonstrate long-term sustainability and

b)   Have a process ready for transferral of data in the case that funding ceases.

The DAC Executive Team is preparing a report with recommendations to ensure sustainable long-term funding for Marine DACs.

Overview of existing portals

In summary, there are a number of Government initiatives creating portals to supply marine, atmospheric and terrestrial datasets. The key portals relevant to the marine sector and their overlap are presented in Figure 2.
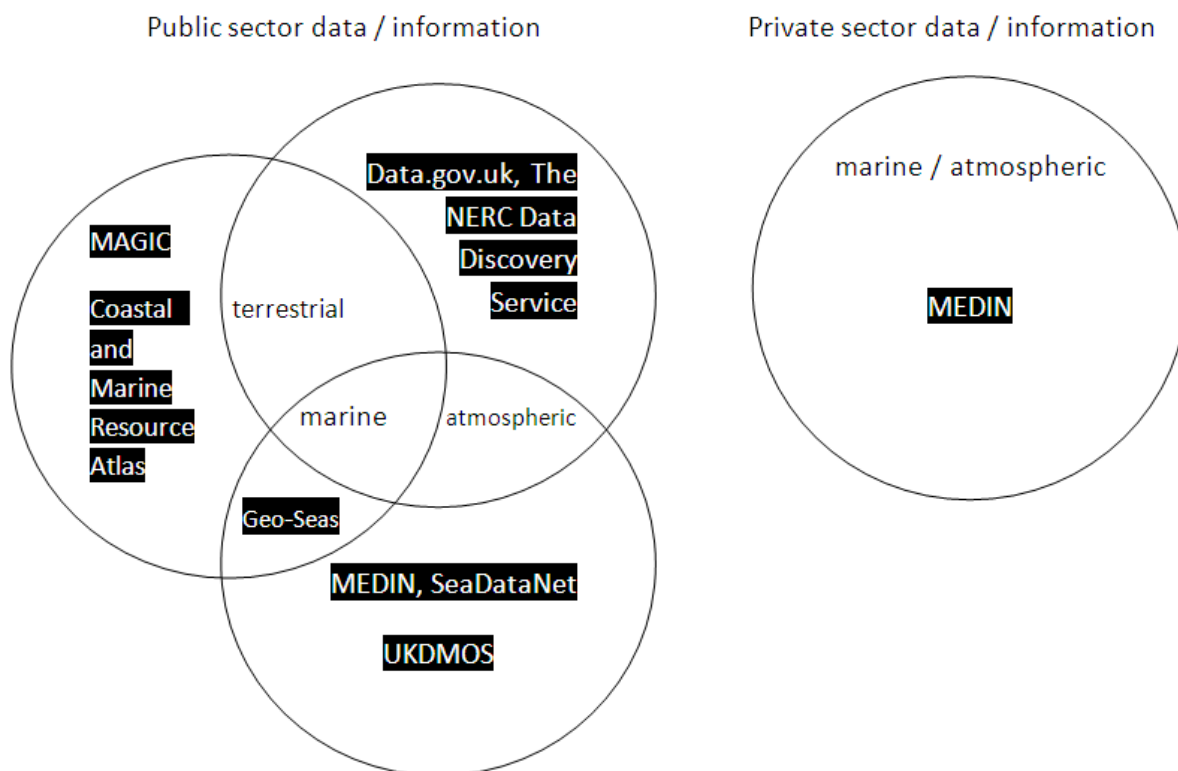
Figure 2. Relationship between marine data / information portals for public and private sector data.

Portal rationalisation

Over the course of the next 5 to 7 years the practicalities of complying with INSPIRE are going to raise their heads. This issue of whether there should be one national portal, which is required according to the wording of the INSPIRE directive, and who should be responsible for that national portal arise. Will it end up being the data.gov portal or a series of portals?

Catering for the wider European public is likely to be problematic. For example, the Met Office is a member of the World Meteorological Organization (WMO) and within the WMO there are regional groups such as the European grouping. A simpler way forward in this case could be to adjust the ECOMET (the European Commission's group of met services) portal so that it will accept everybody's INSPIRE compliant data and metadata.

6.   Issues encountered when trying to comply with guidance

In the final report (MEDIN, 2010) this section presents comments on operational issues encountered by organisations when trying to comply with data sharing and re-use guidance. The comments are from interviewees and are not statements of fact. These include internal issues faced by organisations being compliant and external issues when working with other organisations. The topics covered are:

Internal issues

- Data management
- Data culture

External issues

- Data management
- Formal instructions
- Guidance
- Limitations on access
- Poor agreement of 'standards' across Europe
- Controlled vocabularies for parameter management

- IPR and multiple IPR
- Funding
- INSPIRE element responsibility
- FOI (FOISA in Scotland) and EIR (EIRS in Scotland) requests
- Licensing

6.1. Critical comments about public sector organisations

Comments were made by organisations involved in the audit regarding specific issues that have been encountered and can be seen in the long version of the final report. These comments were noted with the sole intention of flagging perceived problem areas, and have been passed to MEDIN.

7. Conclusions

The conclusions of the data policy guidance review and the data policy audit are presented separately below.

7.1. Data policy guidance review

a)   From the review of guidance relating to charging and re-use of public sector information and discussion with key groups a number of conclusions may be related to MEDIN's objectives. However, the implications of the review do not necessarily need to be actioned through MEDIN, where the role of guidance has the potential to more broadly influence the marine community and marine data sharing policy and re-use. There is extensive guidance available and position statements on Public Sector Information re-use and reviews of the value of commercialisation of marine data. Nevertheless, the situation has been evolving rapidly with data sharing initiatives for Government data through such programmes as Making Public Data Public and INSPIRE requirements, initiatives to release raw and processed data by public bodies and the increasing use of contract clauses with private sector survey companies and developers to grant access to data mean that the situation is increasingly complex.

b)   MEDIN's objectives have been evolving and are perhaps now increasingly aligned to also encompass marine data to meet operational requirements (marine planning etc) as well as promoting data sharing i.e. long-term management of marine datasets; improved access to authoritative marine data held in the network etc.

c)   Whilst MEDIN remains a focus for the marine community data sharing and distributed data management there is potential for wider use through a distributive framework proposed by the 'sharing innovation initiative' of data.gov.uk. The issue remains that many datasets have re-use limitations through IPR and copyright outside the public sector information. MEDIN may focus on comprehensive discovery and search through its portal rather than worry about the re-use and commercial exploitation of data. However, effective recording and presentation of the re-use and

commercial terms becomes important to exploitation, whether charged or freely accessible, and even within simplified licensing regimes (such as Creative Commons type) there is still a demand for attributions.

d)   Where re-use is an issue i.e. the user wants to incorporate the data in other data sets or use the data to develop new techniques and build upon the knowledge base, there will clearly be issues of IPR to resolve. Such potential constraints promote the need for MEDIN (or others) to seek to draw up a set of principles that establish the terms upon which all its participants supply information about data through the MEDIN Portal (and to the DACs). Without such there is danger of a reduced value of environmental information that affects both the data and the value of the DACs themselves. Such principles might seek to modify copyright positions to permit access and re-use on communal terms, including charging policies.

e)   DACs already have a role in data quality assessment and format conversion for archival management, but providing data through the MEDIN portal there will need to be quality assurance built in; otherwise the integrity of such data may be compromised and the original objectives of the whole exercise undermined accordingly. Generally, MEDIN does not engage directly with data accreditation or assurance but rather the MEDIN accreditation process for DACs, provides some measure of quality assurance, although this may not extend to securing quality assurance from within the data suppliers. Although outside the scope of this programme there is a need to address this issue alongside access and re-use.

f)   The reality of marine data managed within DACs is that different categories of data are likely to emerge with differing access constraints, from free access and exploitation through to restrictions for academic use, for consultation and through to commercially confidential data. These differences will largely stem from different IPR and confidentiality requirements, yet the overall objectives of data sharing and knowledge development by the MEDIN community need to be maintained as far as possible. There will be organisations that cannot contribute to liberal access agreement, unless policies change – such as for Trading Funds. Private sector participants might take the same view in order to secure their IPR on behalf of shareholders, although the use of liberal access and publication clauses applied to developers through lease and licence agreements has greatly broadened access – if not re-use terms. Other data may be subject to confidentiality agreements and restrictions of that nature that do not permit full disclosure of data.

g)   For MEDIN to advance access and re-use of data across the marine community following from this review it appears to need:

- to test the boundaries of constraints amongst its members to evaluate how far they would wish to go in IPR negotiations in terms of access, exploitation and charging arrangements for the data in question;

- to evaluate the potential for a more open stance similar to the new 'creative commons' licensing policy announced by OPSI in January 2010; and

- to assess policy related to access to data through the MEDIN portal were licensing and remuneration for rights holders to be provided whilst seeking to develop and maintain the simplest access to the datasets controlled by MEDIN's participant community.

7.2. Data policy audit

From the data policy audit a number of conclusions emerged that relate to both the access to data and metadata, re-use and the charging policies applied. MEDIN in seeking this review has taken a neutral stance as regards charging for data and related policies and primarily is seeking clarification of current understanding and adopted practices. These conclusions are in part being addressed by some of the MEDIN actions or proposals:

a)   Interoperability of data and metadata within and between organisations needs to be improved. Common standards, policies and agreements would help achieve this. This would result in improved efficiency of these organisations when exchanging data and make it easier to work with combined multi-origin datasets. A pan-government agreement (PGA) could push this forward. It was found that there was scope for improving sharing and re-use of data by standardisation of the following; reducing the variety of licences, providing licence templates to simplify data sharing and re-use, and providing contract templates to achieve standardised data generation. For example, reducing dataset projection transformations would save on re-projecting time.

b)   There is variation in data policies between public sector organisations and a lack of standardisation of policy and licence terms and conditions. They are all broadly similar within the business model types, but a lack of a centralised approach was noted with each organisation having typically derived their own documentation. Adopting common licence terms e.g. based on Creative Commons would achieve the same result and is simpler.

c)   The implementation of data policies varies between organisations and whilst the availability of PSI data is improving it requires significant effort to get to the generally expected levels. It was considered that the INPSIRE Directive will improve this situation, but only from the perspective of data access and view rather than necessarily data re-use. There is frustration regarding the lack of detail of organisation's INSPIRE requirements and a practitioner's guidance from UKLP is required, as the implementing agency for INSPIRE. There is still a level of ignorance of the implications and benefits of INSPIRE and the breadth of the data coverage and duty to submit data. Accelerating progress and clarifying strategy were called for.

d)   In the last 2 years, policies regarding the UK's marine and terrestrial environmental data have seen a relaxation of licence terms in some circumstances and there are examples of improved PSI data sharing and re-use (e.g. OS OpenData), however, some examples to the contrary were found. Most notably these were related to the Trading Funds (UKHO and the Met Office). The signs are that in general, access to data is improving and changing significantly, and this ongoing trend is encouraged.

e)   In general, from the discussions held there is a strong desire within the marine data community to share / re-use data when appropriate, although the terms and constraints have been less well articulated. The will is there and the situation is improving with advances in technology helping this situation.

f)   Most public bodies interviewed could improve how they inform the public and private companies as to what data is available for no charge, minimal charges or under a commercial licence. In general, guidance on websites regarding obtaining data, and who to contact regarding it was poor in the case of the larger more disparate organisations.   Some simple measures could improve this situation. A lack of clarity of when this applied to certain data was found to confuse and sometimes frustrate those trying to obtain these data when perceived blockages to data were not

fully explained. There should be clear distinction between products for specific markets and public access to data.

g)   Paying for raw PSI data collected using the public purse does not sit comfortably with many people. Do limitations on data re-use imposed in order to generate revenues outweigh the potential maximum financial benefit to 'UK Plc' by lifting restrictions?

h)   To meet the data requirements of INSPIRE and re-use regulations, significant effort is required to get existing data into the required format. For most organisations this needs to come from existing budgets as no additional funding is available. This situation could increase the time taken to become fully compliant. High quality data which requires months of effort to get it up to the INSPIRE standard could be achieved by Government funding, however, it is more pragmatic that organisations will meet product specifications once these are defined, and make historic datasets INSPIRE compliant only if required to do so.

i)   Data from academia including NERC funded research (which is funded publicly) is not always readily available for sharing and re-use despite grant conditions stating that data must be uploaded to a DAC. The PSI regulations specifically exclude research organisations, thus it largely remains up to data policy of funding organisations to introduce clauses for sharing data, albeit after a period of exclusive use.

j)   There are at least four Government funded data portals which deal with marine data either in existence or being formed at the time of writing (MEDIN, MAGIC, UKDMOS and EOFF). There are many other marine project level portals (MESH, EMECO, MALSF GIS etc). These portals fulfil differing data management requirements, adopting differing standards, access policies and delivering against different policy and legislative drivers. It is not possible to enforce standardisation across these different projects as they meet different objectives and require differing metadata. The important data policy-related issue will initially be promoting greater interoperability to ensure that data searches can search distributed across data stores, but with a longer-term goal of achieving greater harmonisation of standards, if not consolidation. This aim is important in improving access for the user, simplifying accesses constraints and reducing the potential duplications.

k)   The funding of the chosen portal model and associated support must be secure to ensure that a long-term and well-maintained service is available. Present shorter-term funding is a risk to realising such a portal and service. Sustainability and maintainability of data is key, as is a clear strategy. Efforts should be made to ensure that funding is guaranteed. Developing a national marine data policy would help in achieving this.

l)   The 'public use of public data' initiatives providing simpler licence arrangement for certain classes of PSI data through the OPSI and data.gov.uk are seen as broad-scale initiatives that are targeted at IT developer rather than sector users. This initiative seeks to leverage re-use of PSI and reduce barriers to re-use, but is little used by the marine sector. If data.gov.uk holds all of the data there is risk of duplication, contradictory information, out of date data and an inability to meet demand (e.g. during bad weather or volcanic ash events for meteorological data). It is considered that data.gov.uk is unlikely to be able to replace all other metadata portals which address the needs of specific sectors and which maintain both PSI and private sector data, and does not function as a data archive or management organisation. Importantly, sector portals support the detailed functionality and metadata content required by the user community, which is lacking within

data.gov.uk. However, it is important to see how the interaction between these portals and data.gov.uk would operate in practice to benefit the marine data user community.

m)   There is widespread acceptance that improved access to data and metadata are needed to enhance standardisation across the sector and ease re-use. Interoperability and adoption of standards, particularly of metadata and increasingly of data specifications and formats can ease re-use from the technical perspective. Transparent and straightforward documented re-use policies that meet the PSI regulations are a minimum, but more liberal policies and simplified licensing are increasingly seen as leveraging the value in PSI data and meeting public expectations.

n)   Such extensions of the access and charging policies are within the say of public bodies where the PSI does not have third party IPR, but there remain barriers to distribution for private sector data and where IPR restricts distribution under PSI regulations. However, policies around commissioning and licensing of seabed activities provide a potential mechanism that has been used widely in the last few years to open access to otherwise closed archives.

o)   Improving the marine sector's appreciation of metadata is required. Emphasising the importance of creating metadata when surveying and in post-processing will improve the quality of metadata for the data life cycle. Standardisation of metadata as much as possible to make the requirements more easily understood and instilling the importance of data being interoperable will improve the uptake. The more metadata that can be created when generating data, the less effort will be required by third party organisations such as DACs creating it. It is also noted that DACs will probably need to do some work on supplied metadata to generate standardised forms, but this effort can be minimised.

p)   General guidance and strategic planning for marine environmental data users from a single source such as MEDIN is widely acknowledged as being an important function by the community. It is important to recognise that some organisations do not solely deal with marine environmental data, therefore, consideration needs to be made of wider needs to avoid a fragmented approach. A 'family tree' stating who is doing what and how it links together would be of benefit.

q)   Awareness of the regulations and guidance within organisations varies, largely on the basis of the extent of their engagement with PSI and distribution of information. There is much user ignorance of the distinctions between FOI, EIR and PSI particularly in terms of the distinctions between access and re-use. There is also a lack of clarity of the Information Asset Register (IAR) and the value of these in enabling identification of the available data. Waivers and exclusions to Crown copyright and Exemptions of Public Bodies for charging above the marginal costs are also complex to unravel for any particular dataset or information, especially where there is tradeable information within PSI organisations. In particular, those that were not directly engaged with OPSI IFTS tended not to be aware of it. The scheme is voluntary for non-Crown Trading Funds, but its objectives of accrediting Public Sector bodies in transparent and fair use of public sector information and offers clarity to those using the data services of these organisations. Wider awareness of IFTS (particularly knowing which organisations are members) across the community would be beneficial so that the obligations on members are understood. This will make dealings with IFTS members more informed and ensure that data is shared as required. This aims to provide clarity to the user on the licensing and advice on charging for data, where there are exceptions to marginal cost pricing etc.

r)   The role, activities and work undertaken by the UK Location Programme and OPSI is not fully understood by those who work with public sector data. There is scope for more exposure in general.

For example, being aware that complaints can be made to OPSI if it is thought that an organisation is not complying with IFTS, as well as providing support to make sure that your systems are compliant.

s)   Re-use across Government departments is not within the scope of the PSI regulations. It is generally considered not to be right that different parts of Government are competing with each other regarding data. The public are unlikely to accept Government departments charging each other for data.

t)   There is a lack of clarity regarding trading funds and their raw and derived products. Definitions of what is publicly owned but only certain bodies can use, or raw data which is open and both the public sector Trading Fund and for others to re-use should be made. It is either an open or closed market and at the moment it is viewed as being closed. This is considered to be fundamentally wrong and it probably stems from a lack of clarity or a lack of enforcement around the Trading Fund models.

u)   OS licence restrictions are limiting the re-use of OS derived data.

v)   As a Non-departmental Public Bodies (NDPBs) the EA's pricing structure appears to be significantly different to other NDPBs.

w)   In some organisations there can be an internal issue in terms of getting the full engagement of senior management regarding the importance of metadata and making datasets interoperable, to ensure that this issue is addressed across the organisation.

8.   Recommendations

The recommendations are ranked from 1 to 5 with 1 being low priority and 5 being high. Ranking values are based upon the authors' opinion based on comments during interviews. Those ranked 3 to 1 are in Annex 7 of the final report (MEDIN, 2010). Ranking has only been completed for the data policy audit.

8.1. Access policies

a)   Putting a pan-government marine data plan in place will lever the importance of marine data management. Rank = 5.

b)   Organisations to declare what data they hold is PSI (Public Task) on an asset list with INSPIRE compliant metadata. This might usefully be the extension of the Information Asset Register and co-ordination of information relevant to Freedom of Information Act (2000) Publication Scheme compliance and the Environment Information Regulations. Common standards should be used to present such IAR and related information so that interoperable searches can be generated. If all of an organisation's data is PSI then this should be clearly stated. Equally, other information that is produced (either in raw form or post processed) that is deemed to fall outside these publication or re-use obligations needs to the identified. Raw public data has no intellectual property rights. It is important to clarify who is adding value and what that value is. For example, if data is not PSI, such as MCA Automatic Information System (AIS; ship track and other information) data, then a record is available of the reasoning so that every time this issue is raised the process of investigating it does not need to be repeated. Rank = 5.

c)   Investigation to be completed into the most efficient and cost effective method of improving access to and disseminating marine (and other data theme types). There must be many sector-specific organisations like MEDIN focussing on themed data, is this the most efficient way to work or is the data.gov.uk approach a better model? Coordination of similar activities is paramount. This is probably a UKLP or LWEC task. Rank = 5.

d)   There is a need to make clearer the correct contacts for obtaining data and for data specific records of licence terms where there are restrictions beyond a simple licence condition (i.e. third party rights). Despite metadata records of the data distributor, these may not reflect the current position of the contact unless metadata is updated. Also, if the data is passed to DACs with existing metadata and the DAC then becomes the point of contact, then the original contact details might be lost. In the larger disparate organisations this can be problematic and issues such as time taken to locate the right person, and senior staff not being involved in the data release process when they should have been. This could be done from a web page with a generic contact email. This would also ensure that internal procedures for approval of data release are channelled through the correct individuals. Rank = 5.

e)   The provision of no charge, freely available, quality assured national datasets e.g. bathymetry and wrecks, would be of great value to the public and private sector alike. These 'core geospatial data' have been highlighted by the APPSI as essential components to leverage use of thematic data. The key themes for core geospatial data may be similar to those promoted by the INSPIRE annexes, including oceanographic features, administrative boundaries etc. For example, in some cases, rather than purchasing a SeaZone licence, publically available bathymetry datasets serve their purpose but they are not the UK's official dataset. Also, some public datasets, such as wrecks, are only available from SeaZone which significantly limits usage. UK public bodies and companies should be using UK datasets. It is possible that with more raw data being made available that numerous bodies might process the same data for similar processes but end up with slightly different outputs. This should be coordinated to avoid such an outcome and duplication of efforts. For example, if the MMO requires a seasonal sea surface temperature map based on the last 10 years of BODC records who produces the temperature map? Who is going to maintain it? If multiple organisations require these data layers then collaborative funding could realise it and it then becomes publically available for no charge as opposed to selling that map repeatedly on the justification of the value add. Rank = 5.

f)   When data management personnel change the incoming staff can have different drivers. This results in inconsistency in application of policy and hence the services provided. Increased effort must be made to ensure that policy is clearly defined and is transparent for all involved. A common way of working would ease the flow of data. Rank = 4.

g)   The process for raising issues in obtaining data from public sector organisations should be made clear to the whole community, e.g. AIS data from the MCA. There is need for greater awareness of the OPSI role in supporting unlocking of PSI, or in supporting application of the regulations and the escalation process of the APPSI. Issues relating to the sensitivity of some of the related data could be acknowledged by providing a scaled down but definitive version of these data, with justification as to what could not be provided. Rank = 4.

8.2. Operational approaches

a) Making existing datasets INSPIRE compliant will require significant resources which is a big issue for some organisations. Support, guidance and tools will for expediting this process would be valued. Rank = 5.

b) Many data researchers would prefer as few portals as possible to find and obtain data and formats which are ready for re-use rather than an excel spreadsheet or a pdf for example. This is obviously complex as one user may want completely different products and formats to another. In the absence of such a portal a clearer explanation and interoperation of existing portal initiatives is needed. The coordination and potential consolidation of existing portal initiatives should be considered. This could potentially be undertaken by the UK Marine Science Co-ordination Committee and the UK Location Strategy group. Rank = 5.

c) Make it a contractual requirement for contractors to post data to MEDIN DACs or similar. Rank = 5.

d) To reduce repeat extracting of datasets from an organisation's database, hold standard extracted basic underlying datasets centrally either by the organisation for future dissemination or by a central government portal. This will make more data available to pick up 'off the shelf' without going to the counter and 'asking' for it. From large datasets individuals can then extract what they need. Rank = 4.

e) It would be beneficial for data enquiry web pages to request standard information regarding the user and the intended use of the data. At present they can be basic, therefore, a better structure would help all involved; the requestor would be able to better explain their intended usage and the provider could make a more informed analysis of the query and deal with it appropriately. Recording who and for what purpose data is downloaded can feed back into service provision and improve interfacing, but the use of personal data needs to be notified to the downloader. Rank = 4.

8.3. Best practice initiatives

a) That a single common metadata standard, probably INSPIRE compliant, is used across the sector a far as is practicable. Changing metadata standards once adopted is extremely costly and time consuming. The adoption of MEDIN metadata is a central component of the data submission policy, but needs wider dissemination and support to encourage the wider marine data community to adopt this as a minimum requirement. Rank = 5.

b) Work with the European Commission and its working groups setting metadata and data descriptions and to ensure that they are complementary to standards that members of the WMO already have and use, which would avoid a lot of cost and duplication. A potential solution would be the development of metadata translator software between common standards. Rank = 4.

8.4. Pricing and licensing

a) Pan-government terms to avoid ongoing unrealistic charges for licensing their data to other government organisations. Rank = 4.

b) Marginal costs should be compared across marine public sector organisations by OPSI to see if they are comparable and consistent. The basis for calculating marginal costs should be transparent

and equitable between organisations; the IFTS appears to offer an approach, through its audit methods, to encourage consistent approaches between organisations. The charging for data also may need to consider competition law where a commercial organisation may sell equivalent products. However this issue is only likely to occur for products generated outside public task and for derived products rather than raw data, which are still liable for access under Re-use of Public Sector Information (RPSI) regulations. Rank = 4.

9.   References

HM Treasury, 2000. 'Cross Cutting Review of the Knowledge Economy Review of government information', December 2000. http://webarchive.nationalarchives.gov.uk/+/http://www.hm-treasury.gov.uk/spend_sr00_ccr.htm

HM Treasury, 2009 'Putting the Frontline First: Smarter Government' Cm 7753, December 2009.

Mayo and Steinberg, 2007. 'The Power of Information'. June 2007.

MEDIN, 2010. 'Approaches to data policy in the marine sector'. Final report version 1.0, long version (110 pages inclusive). November 2010.

MRAG et. al., 2009. 'Legal Aspects of Marine Environmental Data, Report for the European Commission, Maritime Affairs, November 2008. http://ec.europa.eu/maritimeaffairs/study_lamed_en.html

UK Location Programme, 2010. 'Position Paper Discovery Metadata Service'. Location Information Interoperability Board, Metadata Working Group. DMS Position Paper 20100518 v1

Annex 1 – The contents of the long version of the final report

Executive Summary
Glossary
1.   Introduction
2.   Scope and purpose
3.   Marine data policy guidance review
3.1.   Overview
3.2.   UK PSI distribution and re-use policies
4.   Data policy audit
4.1.   Methodology
4.2.   Differences in approach to policy
4.3.   Description of data sources and access policies
4.3.1. Data sources
4.3.2. OPSI IFTS members
4.4.   Understanding of best practice within the sector
4.5.   Pricing and licensing of data
4.6.   Markets
4.6.1. Market volumes
4.6.2. Opportunities
5.   Initiatives to improve the sharing and re-use of marine environmental data
6.   Issues encountered when trying to comply with guidance
6.1.   Critical comments about public sector organisations
7.   Conclusions
7.1.   Data policy guidance review
7.2.   Data policy audit
8.   Recommendations
8.1.   Access policies
8.2.   Operational approaches
8.3.   Best practice initiatives
8.4.   Pricing and licensing
9.   References
Annex 1 – The regulatory framework and its interpretation
Annex 2 – Statement of Work
Annex 3 – Source material for the marine data policy guidance review
Annex 4 – Interview questionnaire
Annex 5 – Understanding of best practice within the sector
Annex 6 – General marine data working notes and other issues
Annex 7 – Recommendations ranked 3 to 1
Annex 8 – TNA (OPSI) Overview