# Marine Biodiversity Data Flow in the UK


## Dan Lear & Becky Seeley

## The Marine Biological Association


March 2011


**This report was produced The Marine Biological Association of the United Kingdom under the Defra/NBN Trust contract 2008-2011 to improve data provision, management and coordination in the National Biodiversity Network**


**Reference:**

Lear, D.B. & Seeley, B., 2011. Marine Biodiversity Data Flow in the UK. *Report to the National Biodiversity Network*.



# Recording - Marine Biodiversity Data Flow in the UK

## Introduction

The report provides a review of the current level of exchange in marine life data[1] and its management in the UK taking into account the current structures that are in place between data providers, custodians and managers. In addition, the report makes recommendations on how data flow can be improved over the next few years to achieve greater exchange and interoperability within the marine sector.

## Background

Marine life recording has a strong tradition in the United Kingdom and marine data is now collected by a broad range of organisations and individuals. Much of the early work promoting the national recording of marine species was carried out by the Marine Conservation Society (MCS) which published the first edition of Marine Recording in the winter of 1986. Structured recording schemes had been in existence for many years prior to the work of the MCS, including the Norman Holme scheme for the recording of benthic macrofauna and the guidance produced by the Biological Records Centre, both dating from the mid 1970's. Detailed but localised recording endeavours were also available, such as the Plymouth Marine Fauna (1904) or the Fauna of the Clyde Sea (1960's). Many universities with a marine science department developed record card collections in the 1920's and 1930's of species recorded from the area from field other research trips, e.g. North Wales Marine Fauna (Records spanning 1840-1990). Marine recording however has never had the same exposure and numbers as terrestrial recording. Within organisations marine biodiversity data is often collected for specific purposes, often as one off surveys. As one off operations, the data is often then mothballed. Efforts over the last 10 to 15 years from organisations including the National Biodiversity Network (NBN), the Marine Life Information Network (MarLIN),The Archive for Marine Species and Habitats (DASSH) and the Marine Environmental Data and Information Network (MEDIN) have sought to release much of that data, but it is recognised there is still a lack of true data flow.

By simplifying the flow of data between organisations costs for the management and dissemination of data are reduced, and the integration of disparate datasets simplified. In addition, adherence to data flow guidance can reduce unnecessary duplication of data and put critical validation[2] and verification[3] stages into the data

---

[1] Including data from strandline to deep water, collected from professional and amateur sources.

[2] Validation – the process of checking if something satisfies a certain criteria.

[3] Verification – confirmation or additional proof that something that was believed (a fact, hypothesis or theory) is correct.

lifecycle model. Previous work undertaken by the NBN, and in consultation with users and providers of data to the NBN, has produced a series of data flow principles:

**Principle 1:**
Wildlife data flow should enable the production of a single master version of a record promptly, to be maintained by an agreed custodian responsible for sharing the complete record across the National Biodiversity Network with onward access and re-use not restricted without clear justification.

**Principle 2:**
Wildlife data flow should employ efficient management processes which assure the completeness and factual accuracy of records in terms of the identifications, observational location, ownership and permissions to share and re-use.

**Principle 3:**
Wildlife data flow should process records as close to the time and place of observation as possible. Clearly defined roles and responsibilities are carried out by those most familiar with the circumstance, whilst ensuring best use is made of any necessary expertise available at the local, national or international level.

**Principle 4:**
Wildlife data flow should make the intended route and process for onward supply and re-use of records clear, giving the community confidence that they will be managed effectively, shared responsibly and be available to those that need them across the National Biodiversity Network.

One of the primary resources for accessing biodiversity information is the UK National Biodiversity Network, established in 2000 with the prototype NBN Gateway being launched in 2001. The NBN Gateway holds in excess of 61.5 million species records fed through national schemes and societies, Local Record Centres and contributing organisations. The NBN has developed close links with many terrestrial national schemes and societies and Local Record Centres to ensure the flow and dissemination of data. For marine biodiversity data there is not the same number and tradition of recording schemes and data are not collected and stored in the same manner as terrestrial; taxa-focused groups (e.g. Butterfly Conservation Society). Of the many data suppliers to the NBN Gateway, only about 35% hold marine data.

The NBN has worked closely with the Marine Life Information Network (MarLIN), established in 1998 at the Marine Biological Association. MarLIN provides a focus for marine recording throughout the UK. Through the Sealife Signpost scheme, MarLIN provides a portal directing recorders to the most relevant species-focused recording schemes, in addition to establishing a general marine life recording web application. MarLIN is recognised as the marine node of the NBN, and routinely provides survey and sightings data to the NBN that would normally be retained by

individuals or small groups. The MBA has also had a role in collating marine life data from other organisations and individuals and now hosts the Archive for Marine Species and Habitats Data, DASSH. DASSH operates within the framework of the Marine Environmental Data and Information Network (MEDIN), and is an accredited MEDIN Data Archive Centre (DAC)[4].

MarLIN hosts a very successful recording scheme that allows UK-wide records of all marine species to be submitted to the MarLIN Web site; this data is then quality assured by the MarLIN team and progressed to the NBN. Other national marine organisations such as the Marine Conservation Society, Shark Trust, Seasearch, British Phycological Society, Porcupine Marine Natural History Society, Whale and Dolphin Conservation Society and the National Marine Aquarium have established recording schemes with varying levels of engagement with the NBN. In addition to national schemes there are also a number of regional and local groups collating marine biological data as well as some records collated by Local Records Centres. The level of engagement by LRCs with coastal coverage, varies widely and is subject to differing levels of marine data management, often based on the interests and capabilities of existing LRC staff. The recent Marine and Coastal Access Act (2009), the Marine (Scotland) Act and the reporting requirements of the Marine Strategy Framework Directive (MSFD), have highlighted the requirements for quality assured marine data to underpin legislation, planning and the MCZ designation process. As a result there has been an increase in interest from Local Authorities, Statutory Agencies and conservation bodies in marine data. The data needs of the MCZ projects have been broadly met by national data collation exercises such as the MB0102 Biophysical Data Layers contract led by Defra. However, in addition the regional projects have directly sourced data from local stakeholders who may have been unwilling or unable to be involved in a national project. The ability of the regional MCZ projects to engage with smaller groups and individuals is highly commendable. However it is important that the data is shared with national initiatives and undergoes rigorous verification to ensure any contentious decisions are transparent and supported by scientifically robust data.

One major change resulting from the Marine and Coastal Access Act (2009) has been the formation of the Marine Management Organisation (MMO) for England. The MMO is an executive non-departmental public body (NDPB) which incorporates the work of the Marine and Fisheries Agency (MFA) and has acquired several important new roles, principally marine-related powers and specific functions that were

[4] A DAC within the MEDIN framework provides secure long-term storage for electronic data in agreed and open formats. The DAC network (British Geological Survey (BGS), British Oceanographic Data Centre (BODC), UK Hydrographic Office (UKHO), the Met Office and DASSH) provides the capability to upload and retrieve data, and allows data contributors free access to their data managed within the DAC framework. The DAC Accreditation process and criteria can be found at http://www.oceannet.org/library/work_stream_documents/documents/medin_dac_accred_process.doc

previously undertaken by the Department of Energy and Climate Change (DECC) and the Department for Transport (DfT).

The Marine (Scotland) Act is the equivalent legislation in Scotland and is primarily implemented by Marine Scotland, in Wales the Marine and Coastal Access Act falls within the remit of the Welsh Assembly Government (WAG).The data needs of the MMO are still being defined, however, the development of effective marine plans that will be administered by the MMO will hinge upon the provision and availability of high-quality data with a broad spatial coverage. The MMO is currently establishing data sharing agreements with a number of data providers and holders including The Crown Estate, statutory agencies and MEDIN partners and archives. The MMO has recently become a full sponsor of MEDIN and now sit on the sponsors board. The MMO will work with MEDIN and the Data Archive Centre structures to access its requirement for high quality data.There are a number of mechanisms for the transmission of marine data to the NBN, all of which are utilised by different organizations and to differing degrees, software solutions such as Marine Recorder, developed by Exegesis Spatial Data Management (ESDM) and Recorder developed by Dorset Software Services Ltd. In addition the NBN can receive data in the standard NBN Data Exchange Format (DEF), a flat file of relevant information that can easily be ingested by the NBN Gateway.

## Current situation

There are a number of routes via which marine biodiversity data is provided to the NBN, by a wide variety of organisations. In theory, data collected by the countryside agencies (Natural England (NE), Scottish Natural Heritage (SNH), Countryside Council for Wales (CCW) and the Northern Ireland Environment Agency (NIEA)) should flow directly to the NBN, or via collation by JNCC. However even within the agencies there are differing levels of engagement with the NBN. NE are currently collating a full catalogue of their marine data holdings from both national and regional offices which will identify missing datasets to send to the NBN and MEDIN DACs. Most SNH point data and some NE data are available currently on the NBN via the JNCC database. CCW and NIEA routinely enter their marine biodiversity data into the Marine Recorder application and the majority is now available through the NBN Gateway.

Larger organizations including Cefas, Marine Scotland and the Environment Agency have internal data management structures that meet their operational requirements. However, there is a need for significant resources to be allocated to the task of aligning the internal data operations of such large organizations with the national, harmonized approaches to data management. Much progress has been made already, as marine Scotland and Cefas have commenced work on the generation of discovery metadata records for their data holdings that will be made available through the MEDIN portal.

In addition many individual recorders provide data to taxa-specific or regional groups, including national schemes and societies and the Local Record Centres. From these groups the data is then shared directly with either the NBN or via DASSH for archiving and then on to the NBN Gateway for dissemination. Currently there is only a one-way flow of data from DASSH to the NBN, and as such not all data are archived within the MEDIN Data Archive Centre framework. Recent work has taken place to address this gap in data flow, and regular timetabled updates are planned to commence in 2011.

## Marine Data management in the UK

In the terrestrial sector regional biological data is primarily stored in Local Record Centres. In the marine sector the situation is different as the management of marine data in its widest sense (including bathymetry, hydrographic, chemical and biological data) is co-ordinated by the Marine Environmental Data and Information Network (MEDIN). MEDIN provides the framework for the harmonisation of marine data, and any future developments should take account of the MEDIN model of standards and specifications, underpinned by a series of centres of excellence, the network of Data Archive Centres (DACs). MEDIN provides a mechanism for UK Government and their agencies to meet their obligations under the EU INSPIRE (Infrastructure for Spatial Information in the European Community) Directive (Directive 2007/2/EC of the European Parliament) within the marine sector, and make data available to the UK Location Programme. The UK Location Programme aims to improve re-use and sharing of all public sector information in the UK. The Scottish Government, Northern Ireland Assembly and Welsh Assembly Government have yet to clarify their plans for meeting their INSPIRE obligations.

MEDIN provides and promotes the standards and specifications required to streamline the management of marine data. MEDIN and its partners have developed an INSPIRE and GEMINI 2 compliant discovery metadata standard and a number of data guidelines for the collection and supply of marine data (see www.oceannet.org for more details). The adoption of these standards ensures that the ingestion of data by the MEDIN DACs and end-users is simplified; reducing costs and improving interoperability.

Operating within the MEDIN framework and acting as the MEDIN Biodiversity DAC, DASSH now regularly updates the NBN Gateway with the biodiversity datasets it hosts. In addition DASSH has integrated NBN dissemination into its data sharing agreements and with those organisations forming the proposed distributed MEDIN fisheries DAC (Cefas, Marine Science Scotland and AFBI) to ensure that relevant data is shared with the NBN.

In any discussion on data flow it is important to define the difference between the dissemination and archiving of data, and how these roles are reflected by MEDIN DACs and the NBN. It was agreed at a trilateral NBN/DASSH/Defra meeting in 2007 that DASSH should operate as the archive for marine biodiversity data and the

NBN Gateway should act as a key dissemination route. The agreement has been promoted to a variety of groups, but it is clear that these statements and definitions need to be more effectively and more widely communicated.

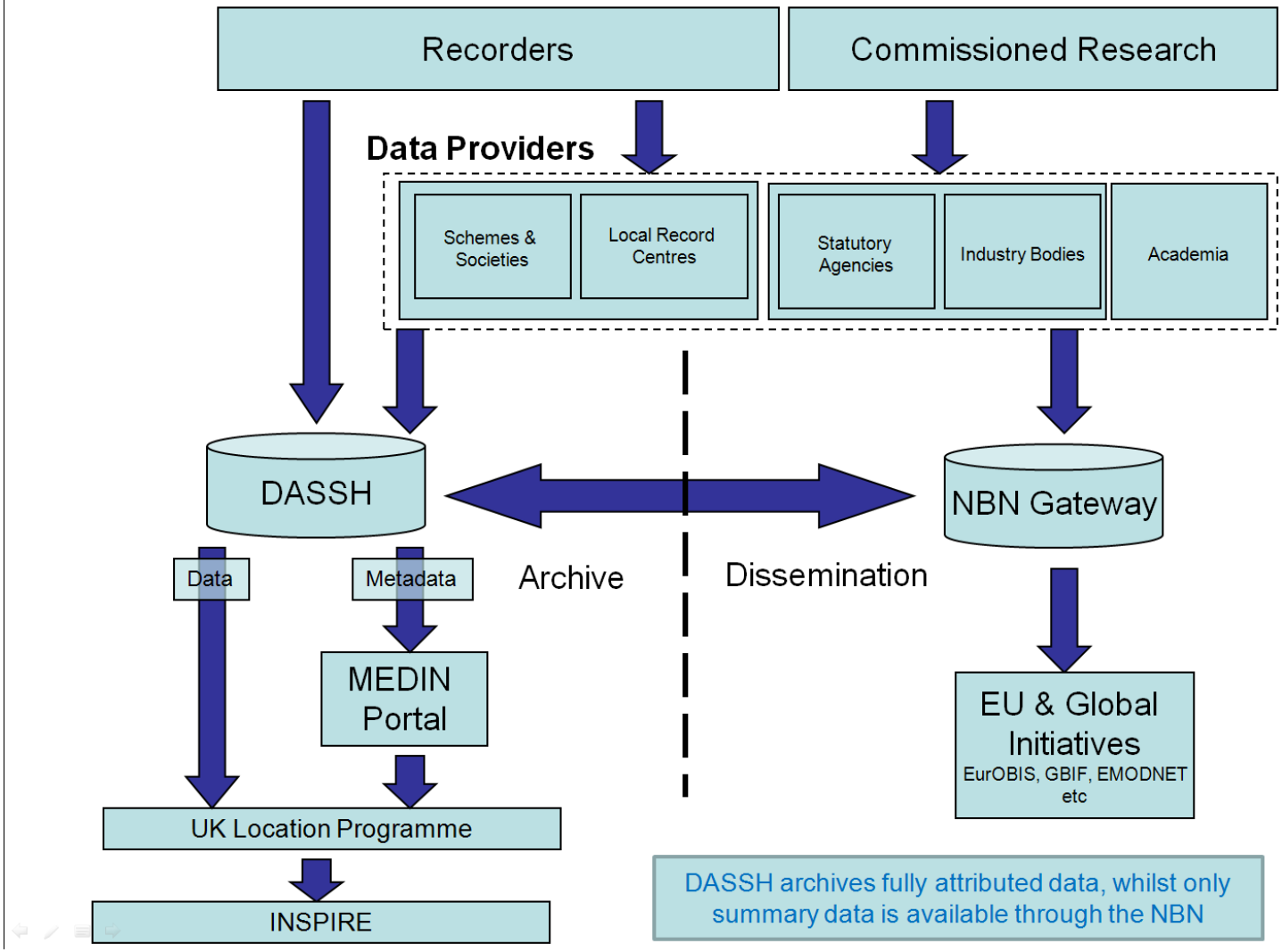


*Figure 1. Recommended data flow within the UK*

## Issues

A number of issues exist across the marine recording community in the UK. In general the marine recording community is small with many groups and individuals facing technical barriers in the management of their records. In addition there exists a certain degree of confusion amongst recorders over where to submit their records. This confusion stems from existence of a number of projects and organisations offering to store marine biological data.

In recent years there have been a number of data collation exercises that have included the requirement to lodge data with an accredited MEDIN DAC, which has improved data flow, but these projects have been those initiated by Defra, Marine Scotland and the Countryside Agencies. The wider inclusion of data archiving in contract clauses and expected deliverables is still not commonplace.

In many cases short-term or species-specific recording programmes are not consistent, with some recording schemes running for a few months annually or even less frequently. Those organizations involved often do not have sufficient resources or expertise to engage fully with the NBN or MEDIN DACs and so the data is unavailable to the wider community.

Data sensitivity is an important issue in the collation of biodiversity data. Datasets may be sensitive because they are of commercial or academic value and therefore more rigorous controls on dissemination are required. Some survey information such as a location name or a surveyor name may be sensitive and require masking of names and reduction in the spatial resolution of a record. Some species are also considered sensitive to exploitation or harassment and records of these species are required by some providers only to be publically shared at a coarse resolution, e.g. 10km square, to protect the species from exploitation. This may only include specific stages of a lifecycle e.g. seal haul out site or female cetaceans with calves.

Data collected by the private sector is often unavailable due to commercial drivers, particularly in the oil & gas, water and renewables sector. However recent changes to Petroleum Operations Notices issued by DECC mandate the engagement with MEDIN. In many cases these data are collected by consultancies, which do not own the data and are not resourced to manage or disseminate it after completion of the contract. The mechanisms for the sharing of data from the commercial sector differs between sector, consultancy and even within teams; the development of the MEDIN guidelines, providing they are simple to adopt and implement should mitigate this issue.

Fisheries data held by Cefas, Marine Scotland, the Agri-Food and Biosciences Institute (AFBI) and Inshore Fisheries and Conservation Authorities (IFCA's – formally Sea Fisheries Committees (SFC's) for English and Welsh waters) has been historically difficult to access however significant progress has been made in recent

years. The organizations are all committed to the development of MEDIN Data Archive Centre for the management and dissemination of fisheries data.

The legacy data held by many organizations dates back several decades in some cases. The costs associated with digitization, quality assurance and mobilisation of such datasets are significant, and require targeted projects to ensure the datasets are secure and available for future reuse.

The legacy of devolution presents additional challenges with the devolved administrations requiring greater control and access to data relating to their administrative areas. Unless standards and structures are agreed and adhered to at all levels there is a risk of a proliferation of national data silos emerging, without consideration for interoperability at the UK and European scale.

The current financial climate and cuts in public sector funding represent a threat to the continued support and management of data provided by volunteer and amateur recorders. Currently many organizations assist through the development of standards, protocols, guidance and technical help. Significantly reduced funding will impinge on this capability, and could limit valuable data being collated as part of the national resource.

## Future Developments

If the UK is to have a cohesive network of Data Archive Centres, as recommended in the Cowling report on marine data (Cowling, 2005) and championed by MEDIN, then clear lines of data flow must be established, including dissemination via the NBN and the archiving of fully attributed data with an accredited MEDIN DAC. As shown in Figure 1 there are a wide range of data providers within the marine sector, and many have existing preferred pathways for submitting these records.

The generation of large volumes of data is now much simpler thanks to crowd-sourcing initiatives such as OPAL, iSpot and the numerous Bioblitz events that have taken place in recent years. The increase in records resulting from citizen science, make the clear flow, validation and verification of records from these sources more important than ever.

By improving the linkages between the NBN and DASSH, and ensuring the tools are available to export and ingest data and metadata in the standard MEDIN format, the barriers to data exchange are greatly reduced.

## Recommended Data Flow Model

Generally data flow can be improved and simplified by the widespread adoption of standards and guidelines such as those developed and promoted by MEDIN. The guidelines provide a basis for the exchange of data in a defined form allowing the development of tools and applications to facilitate the aggregation and integration of disparate data; normally a resource intensive task. However even with the optimum situation of universal adoption of MEDIN standards and guidelines for data transfer

there remains a huge volume of legacy data. DASSH has commenced the work of harmonising access to such legacy data in a standard data model, but the task is ongoing as one of the core operations of DASSH as a MEDIN DAC.

Figure 1 provides a schematic of an idealised data flow model for the marine sector. Whilst the NBN Gateway currently only displays marine species information, the provision of habitat polygons has recently been implemented for terrestrial habitats, with the provision of marine habitat polygons planned in the future.

It is important to note that the NBN Gateway provides a route for dissemination back to the Statutory Agencies, National Schemes and Societies and other data providers. This loop is excluded from the diagram for the purpose of clarity. Summarised data from the NBN Gateway can be shared through the facilities provided by the NBN Gateway and the NBN Web Services, and can be facilitated by the Data Exchange Agreement developed by the NBN Trust. Additionally DASSH is developing Memoranda of Understandings with a number of key organizations to ensure their data is archived in-line with MEDIN recommendations.

Currently DASSH only provides metadata to the MEDIN Data Discovery Portal (http://portal.oceannet.org/search/full); however in time the MEDIN DACs will provide view and download services compliant with the EU INSPIRE Directive for access to full, attributed data through the UK Location Programme framework. The UK Location Programme (UKLP) is working in cooperation with data.gov.uk to establish a central national framework to support UK organisations in meeting the INSPIRE obligations and is based on the concepts of "Data Providers" and "Data Publishers". Figure 2 shows how MEDIN will act as a Metadata and Data Publisher for the marine community. More information on the UK Location Programme can be located at http://location.defra.gov.uk/ .

The Data Flow Diagram in Figure 1 demonstrates the multiple paths available to data providers. However, it is important that records and datasets are submitted to a single collating or publishing organization. The availability of multiple receiving organizations, provides an inherent flexibility and recognizes that providers of marine biological data often have existing relationships and pathways established over many years. The key to effective data flow is to recognize the pathways and formalize the relationships between the collating organizations, namely the NBN Gateway, DASSH, Local Record Centres and National Schemes and Societies as the focus for professional and amateur collected data.

Whilst not explicit on the data flow diagram it is important to recognize that there are quality control and quality assurance steps at each stage of the data lifecycle. Tools provided by the NBN such as the recent Record Cleaner application and the ability to comment on records on the NBN Gateway provide the capability to validate and verify records which may have been sourced from a wide variety of data providers, with differing skills and expertise. The NBN Record Cleaner tool is a stand-alone application allowing the validation and verification of datasets held in

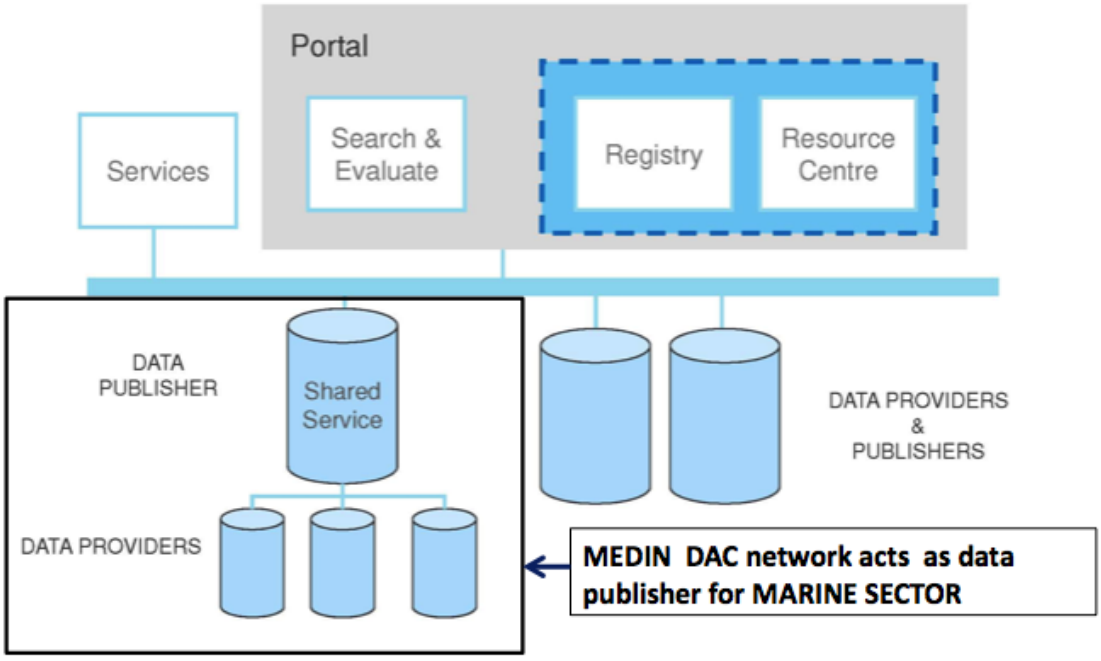


*Figure 2. MEDIN as a thematic metadata /data assembly and publication centre for the marine community, into the UKLP framework*

text files, spreadsheets, and the Recorder software, amongst others. The tool utilises a number of rules to assess the data and allows validation of attributes such as dates and spatial information and verification checks against known distributions and potential identification problems. The Marine Biological Association is currently producing rule-sets for the NBN Record Cleaner tool for in excess of 450 marine species.

Data submitted to DASSH are also subject to checks including those against the known distribution of a species, its ease of identification and the timing of the sighting in relation to known lifecycle and migratory patterns. In addition the validity of the data and associated attributes are checked with any queries being related directly back to the data provider.

## Recommendations to improve data flow to the NBN

1. Ensure that metadata can be imported and exported from the NBN Gateway and Marine Recorder to the MEDIN discovery standard including compliance with INSPIRE and UK GEMINI2. Currently the NBN metadata is not MEDIN compliant and requires some additional fields to meet the requirement.
2. Improved marine data dissemination. Currently DASSH uses the NBN to disseminate as much data as it is able to. However some improvements to the NBN Gateway would improve the NBN's data provision, including;

- Ability to download absence data. Currently absence records are stripped from the records before display on the NBN Gateway.
- Ability to search by location. Many data requests are location specific, in response to surveys or planning enquiries. The 10km grid used widely in

terrestrial systems is not applicable to marine systems, additionally the 10km grid presents a barrier to integration of data from the Channel Islands and towards the offshore limits of the UK Continental Shelf. This functionality is planned for 2011, including the ability to search by the new MCZs, (when finalised).

- Ability to select records by survey, in addition to species, site, dataset or 10km grid square.
- GIS friendly services - Provide services, tools and support to enable data users to integrate Gateway data into their own systems.
- The provision of a chart backdrop to the Gateway for the viewing of marine records would provide greater context and utility once the records move away from the coastal zone
- Ability to map records in WGS84 as the OSGB spatial reference system distorts offshore areas where a number of key species and habitats including cetaceans, cold water corals, deep sea fish and deep-sea sponge aggregations occur.
- Investment in appropriate data structures, guidelines and facilities to manage data flow and provide tools and guidance for those organizations unable to fully engage due to technical or resource based difficulties. Suppliers of marine data should not have to provide their data more than once. The responsibility is with the collating organizations to promote the data flow process, to share the data with permission from the data owner and to ensure there is no duplication between holdings.

3. By data providers giving download access to their data on the NBN Gateway and/or authorisation to DASSH to publish their data on the NBN gateway on their behalf, it is possible for European and worldwide initiatives, such as GBIF to share the data and for data providers to contribute to a wider resource.

## Conclusions

Whilst data flow for marine data is still far from fully streamlined in the UK, huge in-roads have been made in recent years. The work of MEDIN has meant that the marine sector is now at the forefront of developments in geospatial metadata creation and data sharing, rather than trailing behind its terrestrial counterparts . The development of the MEDIN Discovery Metadata Standard (INSPIRE and GEMINI 2 compliant), and the data guidelines will ultimately lead to standard formats for the exchange and storage of marine data. The wide adoption of these guidelines will be critical to removing barriers to data flow.

It is also important to retain flexibility and choice for individuals and small recording groups in where they lodge their data (directly to the NBN or DASSH, or via a Local Record Centre or National Scheme or Society). Without such flexibility it is likely that participation will be discouraged and effectively installs a barrier to the release of data from a vital source. However this does not mean that records should be submitted to multiple organizations. Whilst checks are in place to remove

duplicate records, this can be resource intensive operation for those organizations involved in the collation and aggregation of records from a variety of sources.

The development of exchange tools that can read and write to the standards and templates for metadata and data promoted and developed by MEDIN and the DACs will improve the linkages and reduce the resources required to archive and disseminate marine biological data to the widest possible audience.

## References

Cowling, M. (2005) Marine Data, Where to now? A report commissioned by Defra through the Inter-Agency Committee on Marine Science and Technology (IACMST). DETR, London