



Marine Life Information Network (*MarLIN*)

**Access to seabed marine biological data.
Methods for data cataloguing and acquisition.**

Guy Baker

June 1999

Reference:

Baker, G.J. 1999. *Access to seabed marine biological data. Methods for data cataloguing and acquisition.* Report to the Joint Nature Conservation Committee from the Marine Life Information Network (*MarLIN*). Plymouth: Marine Biological Association of the U.K. (*MarLIN* Report No. 3.)

**Access to seabed marine biological data.
Methods for data cataloguing and acquisition.**

Contents:

1. Introduction.....	5
2. Background.....	5
3. Requirements of a metadata catalogue.....	6
3.1 Introduction.....	6
3.2 The metadata collection form	7
3.3 Transferring metadata to NBN and EDMED catalogues	8
3.4 Conclusions for cataloguing metadata	8
4. Trialling data access	8
4.1 Introduction.....	8
4.2 Level of Access	8
4.3 The data provider	9
4.4 Remote access to datasets.....	10
5. Data cataloguing and acquisition: problems and suggested solutions	11
6. Hardware/software issues	11
6.1 Software.....	11
6.2 Location of the server.....	11
7. Conclusions.....	11
8. Acknowledgements	11
Appendix 1: Glossary of acronyms.....	11
Appendix 2: Metadata Collection Form and Guidance notes.	11

Marine Life Information Network (*MarLIN*)

Access to seabed marine biological data. Methods for data cataloguing and acquisition.

Guy Baker

1. Introduction

MarLIN is establishing access to a network of related data sets on marine habitats and species, building on the database developed for JNCC's Marine Nature Conservation Review (MNCR) programme. *MarLIN* aims to catalogue seabed data holdings and to access datasets to develop contextual information on marine habitats, species and communities. This scoping study, funded by the JNCC aims to:

- establish the requirements of a data catalogue;
- identify the electronic platform best suited to access those data holdings, and
- trial access to data using an 'experimental' dataset.

The study is part of a wider project to interlink datasets from a variety of sources so that they can be interrogated collectively over the Internet. Through a *MarLIN* network, it will therefore be possible to access a much wider range of information than is available to any individual organisation. The project to inter-link datasets is a part of the 'Seabed Data Acquisition and Interpretation sub-programme' within *MarLIN*.

2. Background

Various organisations and agencies, public and private, collect and maintain environmental data often at considerable cost. Information on the nature of a data resource, its format, owner, quality, availability, spatial and temporal coverage, should be widely available. Then maximum value can be made of these data to improve environmental decision making, offer opportunities for education and benefit other interested parties.

Standards for cataloguing metadata are being developed by the European Environment Agency (EEA) in its Catalogue of Data Sources (CDS) (<http://www.eea.eu.int/frseain.htm>), and by the European Directory of Marine Environmental Data (EDMED) (<http://www.nbi.ac.uk/bodc/edmed.html>) co-ordinated by the British Oceanographic Data Centre (BODC) at Bidston. The National Biodiversity Network (NBN) has designed its index to take account of existing structures used by its partners and to adopt emerging international standards, in particular those of the European Environment Agency's CDS. *MarLIN* will contribute to the NBN through development of the marine focus.

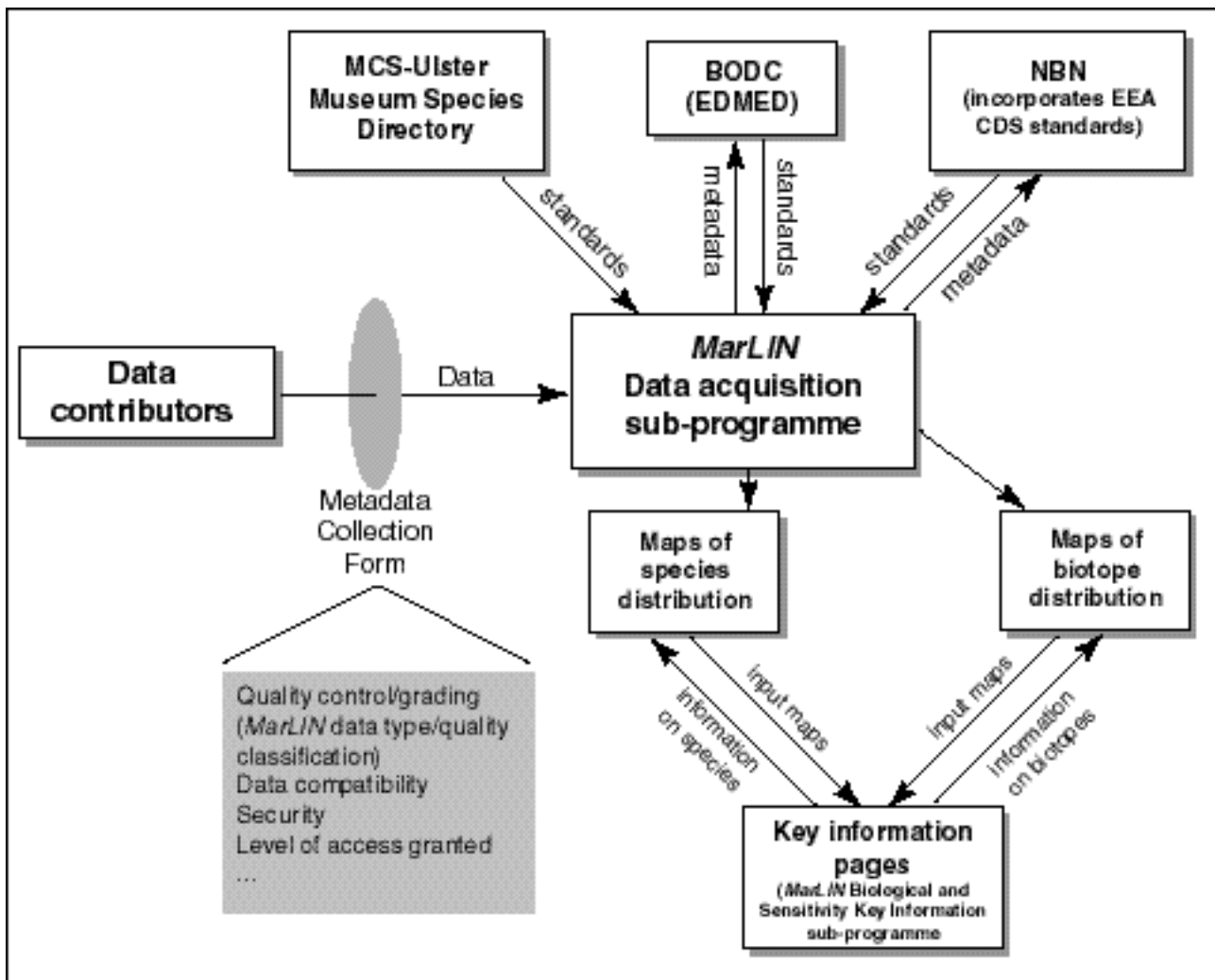


Figure 1. MarLIN metadata catalogue and seabed data holdings: key relationships

MarLIN is working closely with major organisations holding data to access specifically seabed biological data holdings. Using the raw data accessed from various data providers *MarLIN* aims to develop contextual information about marine species, habitats and communities in response to a wide variety of queries. Many issues are raised when accessing existing field survey data. They include the level of access granted by the data provider, ensuring data security, quality control, what format the data are held in. A further phase of the project to inter-link datasets involves demonstrating the feasibility of accessing a trial dataset and a MNCR database simultaneously and seamlessly over the Internet.

3. Requirements of a metadata catalogue

3.1 Introduction

MarLIN aims to generate information on species distribution and to map biotopes. The emphasis is therefore on cataloguing datasets that contain geo-referenced benthic species data. Datasets that *MarLIN* will access are likely to range from fully quantitative data to qualitative data from casual records. The metadata catalogue will have to accommodate that range. The *MarLIN* data classification and definitions of data quality are given below.

Data Type	Description
Fully quantitative data.	All sampled taxa identified to distinctive entities including c. 80%+ to species. (This category usually applies only to samples collected from sediments where organisms occur as countable individuals. Where samples are sieved, 0.5 mm or 1 mm mesh will have been used and quality control measures taken.)
Semi-quantitative but comprehensive data.	Expressed as defined abundances or rough counts or % cover but for all conspicuous species. (All data which includes a mixture of solitary and colonial/crustose species.)
Semi-quantitative data from checklist of species.	
Semi-quantitative data from casual records.	
Qualitative but comprehensive data.	(Presence-absence data from surveys where all sampled taxa were identified to distinctive entities including c. 80%+ to species.)
Qualitative data from checklist of species.	
Qualitative data from casual records.	
MarLIN quality measures	Description
Good.	The data has been collected by a reputable institution with high accuracy standards. (This might be a laboratory which is part of the National AQC scheme)
Adequate.	The data is reliable but may include minor inaccuracies in species identification which would be easily spotted.
Limited value.	There are errors in species identification and or quantification which might not matter if species are anyway widely distributed.
Best guess / dubious.	There are most likely to be errors in species identification and or quantification which matter but the records are to be kept in case unusual records prove to be of historical importance.

3.2 The metadata collection form

To try and avoid duplication of effort *MarLIN* has been looking to existing standards for cataloguing metadata. The NBN metadata standard was itself derived from an extensive review of metadata cataloguing schemes used throughout the world. It is implemented in Microsoft Access (97) and is supported by a data entry application that consists of a suite of Access databases linked to NBN standard dictionaries for taxa, biotopes and administrative areas, and to the GEMET (General European Multilingual Environment Thesaurus) keyword system. (at: <http://www.mu.niedersachsen.de/cds/documents/ws0996/gemet02.html>)

The NBN describes GEMET as being developed by the European Environment Agency and providing a common reference between the NBN Index and the EEA's Catalogue of Data Sources. The NBN dictionaries available via the data entry application help to standardize data input and allow a standardized means of referencing by species or habitat/biotope. (The NBN index can be found on the World Wide Web at <http://www.nmw.ac.uk/nbn/>.)

BODC's EDMED has a less structured free text format that allows data providers to describe their resources in their own terms, and also facilitates gathering of metadata electronically or on paper. EDMED is accessible on the World Wide Web at URL <http://www.nbi.ac.uk/bodc/edmed.html>

The *MarLIN* metadata collection form is based upon fields in the NBN index and EDMED, and has several extra fields reflecting the specific requirements of *MarLIN*. It is intended that *MarLIN* staff will visit data providers and enter the data onto the form, thus saving time and assuring data providers that their information is of value and will be used. The form and its Guidance Notes are in Appendix 1.

3.3 Transferring metadata to NBN and EDMED catalogues

The *MarLIN* metadata collection form is designed so that all relevant information about the data resource is captured and the fields correspond as closely as possible to those within the destination databases. As far as possible, field names and sizes comply with those used by the NBN since this is the more structured catalogue, and any metadata appended automatically to it needs to have the same format. In contrast the free text format of EDMED has minimal format restrictions and should allow input of data from analogous *MarLIN* fields (L. Rickards, pers. comm.). By liaising with partners in this pilot study (JNCC, BODC, NBN and the Environment Agency) from the outset it is hoped that data migration problems can be minimized. When metadata collection from a wide range of sources begins it is envisaged that an electronic form (probably implemented in Microsoft Access) will be used to gather metadata.

Feedback from partners suggests that contact with data providers is important. People are more likely to talk to someone for an hour than to spend 15 minutes on a form that has been widely distributed.

3.4 Conclusions for cataloguing metadata

Cataloguing metadata should:

- adopt existing or emerging standards;
- use standard lists or keywords to facilitate searches, and
- make forms simple and brief or presented in a way that minimizes time needed for completion.

4. Trialling data access

4.1 Introduction

A further phase of the current project required a demonstration of simultaneous access to a trial dataset and MNCR data over the Internet, using software under development by JNCC. In order to mesh with this system the dataset used in this scoping study was required to:

- be in the public domain;
- include species records, and
- have a format that allows a site specific report to be produced (i.e. be geo-referenced).

4.2 Level of Access

MarLIN will use data provided to develop contextual information on species and biotopes and does not make raw data available to third parties. Any organisation or individual contributing data to *MarLIN* retains ownership and copyright of the data, and would make clear the level of access it was prepared to give to third parties. In a meeting on 18 March 1999 to consider the overall philosophy and framework within which a *MarLIN* network of shared data might operate, the *MarLIN* Data Access Sub-Group established two ways by which a data provider could restrict access:

- a. by reducing the amount of information released (releasing broad scale information but not the detailed raw data collected at each station), and

- b. by reducing resolution of data by specifying the scale at which they could be used, i.e. releasing survey locations to a precise six-figure grid reference provided it is only used to the nearest 1x1 km square.

4.3 The data provider

The Environment Agency's National Centre for Environmental Data and Surveillance (NCEDS) holds National Marine Monitoring Programme (NMMP) data, and agreed to provide data for the purposes of this study.

The data used in this pilot study were a subset of the NMMP data consisting of all the Environment Agency's marine benthic and associated data from surveys in England and Wales for the years 1993 and 1994.

Benthic macrofaunal samples were collected from 87 estuarine and coastal sites throughout the United Kingdom between 1992 and 1995. Samples were taken as part of the NMMP, initiated to co-ordinate marine monitoring in the UK. These data are collected yearly by the Environment Agency, Centre for Environment, Fisheries and Aquaculture Science, Scottish Environment Protection Agency, Scottish Office Agriculture, Environment and Fisheries Department, Fisheries Research Services Marine Laboratory, Department of Agriculture for Northern Ireland and the Environment and Heritage Service. The NMMP data typically includes quantitative counts of solitary species but only presence-absence records of colonial organisms. It does not include algae.

NMMP data is held on a database called UNICORN developed by Nigel Grist of Unicmarine. The Unicorn database serves two main functions: maintenance of taxonomic information and facilitation of data entry. The underlying structure is a Microsoft Access 97 database consisting of multiple linked tables, with referential integrity (N. Grist, pers. comm.). It has a front-end providing various functions including data entry (e.g. "Add survey") and output. The output options include Access "Report table", "Primer" format and "Custom" (build your own query), Figure 2 shows a generalised scheme of the UNICORN database.

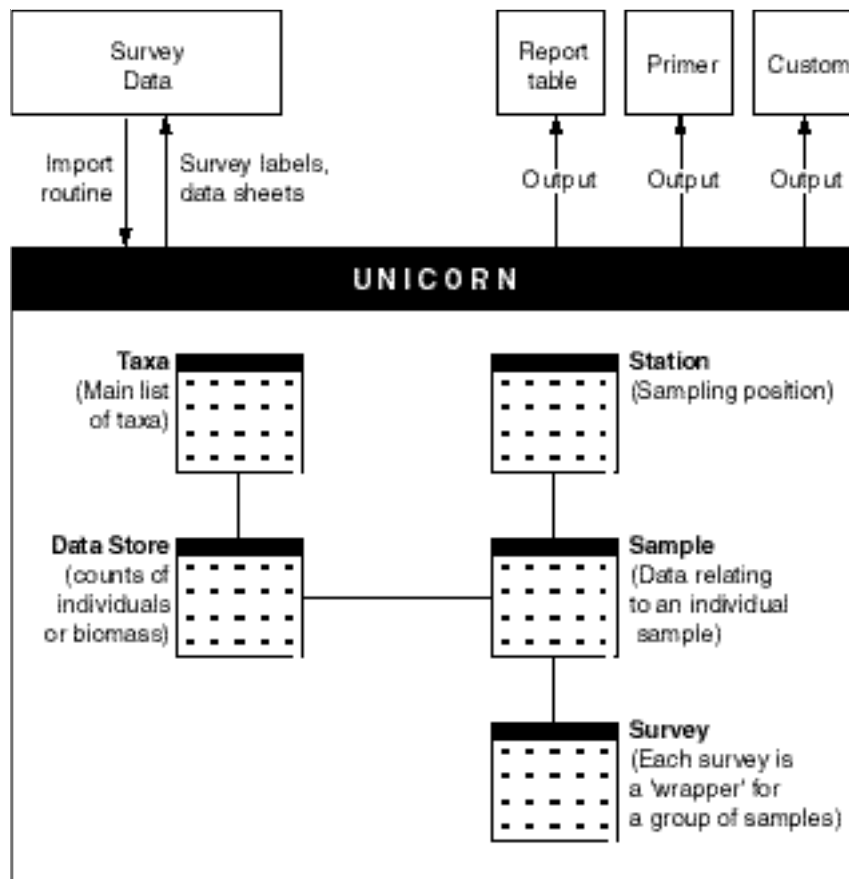


Figure 2. Generalised scheme of the UNICORN database

The dataset under consideration was extracted from UNICORN by a custom Microsoft Access query written by Graeme Storey of the NCEDS in Bath, and was provided in the form of a Microsoft Excel spreadsheet containing a matrix of 1200 species from 39 NMMP monitoring sites (see Appendix II).

4.4 Remote access to datasets

In response to a user request for information, a query would be generated by the *MarLIN* server and extract information from relevant datasets. Reading (and writing to) remote databases via web pages is possible using the main methods for running server-side processing which are Java, ASP, (Active Server Pages) CGI (Common Gateway Interface) or delphi.dll technology. Using these methods the client (browser) is only responsible for submitting a query and displaying the results, and therefore requires only web browsing software. The Web server (*MarLIN* or data provider?) accepts the query, creates a connection to the database server, executes the query, formats the results into HTML (Hypertext Markup Language) and sends the HTML to the client. The relative merits of the four methods depend upon the application.

There are two scenarios that appear likely to arise in response to the question of remote access to datasets:

- a. **A dynamic link between *MarLIN* and the data providers server.** A live link will be achieved using JNCC Java objects to dynamically query remote databases and present the results within *MarLIN* web pages in response to users queries. The data provider is willing to mount the necessary JNCC software on their server. No direct link is made between the browser and the database server, which is an important security safeguard. Additionally, the data provider would probably maintain a read-only copy of the dataset on their server, and it

would be this copy that would be queried via *MarLIN*. Also, queries will need to be standardised through consultation to allay concerns of damage to data.

- b. **Static copies of datasets on MarLIN server.** If the data provider is unwilling to allow the kind of dynamic link access outlined above, it may be necessary for MarLIN to hold a static copy of the dataset that would periodically be updated by a process that would depend on the size of the dataset involved and the frequency with which it is updated. This process could range from delivery of a floppy disk once a year, to an automated FTP routine that was triggered once a month. The latter process would require that an FTP site be set up at the data providers server (a standard feature of any quality server software such as NT server), and configured with password and permissions so that only MarLIN could initiate the download. Having potentially large datasets requiring frequent updating, residing on the MarLIN server is less desirable since the dynamic linking aspect is lost, and the demands on the MarLIN server may be excessive.

The EA do not consider direct access to be the best option because *MarLIN* would not have access to all of the data on the UNICORN database. Also, the trial dataset provided by the Environment Agency is not large (754 KB, a high-density floppy disk has a capacity of 1440 KB) and will be updated yearly. Even if all 87 NMMP monitoring sites are included in subsequent donations, these data will fit on one or two high-density floppy disks. In addition consuming server resources for such a small dataset is unnecessary. Establishing an Internet link to the NCEDS server is considered inappropriate in this case.

Access to datasets over the Internet will require a live link between *MarLIN* and the organisation in question, but agreement must be reached with the organisation about mounting the JNCC objects (see section 6.1) on their server. They must be satisfied that changes will not be made to their systems and that their live data is safeguarded by the procedure used by *MarLIN* to query the dataset.

5. Data cataloguing and acquisition: problems and suggested solutions

Problem	Solution
Data input Some datasets may be in older spreadsheet format or exist only as hard copy. Older, non-relational database packages can also present data export problems.	Specially developed software can extract data from hard copy? Old software?? Dan? Most widely used database and spreadsheet packages allow data export in standard formats. For hard copy records, <i>MarLIN</i> undertakes to input data by hand, if necessary. The dataset used in this pilot study can readily be imported into the database package used by <i>MarLIN</i> .
Spatial location with reference to <i>MarLIN</i> objectives Location may be expressed in very general terms (e.g. "present in Bull Bay") to much more specific locations (e.g. 50° 02.098'N, 05° 03.742'W)	Data will be incorporated at the finest scale possible. The coarser the scale, the more restricted will be its use. For instance "present in Bull Bay" would map on a 10x10 km grid.
Taxonomic Codes Individual records within a database should be uniquely identifiable. At the time of writing there is no standard taxonomic code in use for marine biodiversity projects. This	It appears that the ERMS is emerging as a standard, and as such it should be incorporated in the <i>MarLIN</i> database when ready. In the meantime, the MCS/Ulster Museum directory is

<p>makes cross-referencing between datasets on the basis of species names problematic. Taxonomic codes in use or development are given as a footnote to this table:</p>	<p>most widely used and codings should be adopted by <i>MarLIN</i> (although this requires electronic access which is currently not available).</p>
<p>Species abundance The majority of species collected during remote sampling programmes are stored as a number (count per unit area), however certain species, e.g. colonial animals may be recorded as present/absent. For biotope definition, presence/absence information is of limited use. Quantification of distribution information is desirable.</p>	<p>Presence/absence data should be identified as such and used for mapping species distributions. An 'algorithm' is required to identify the MNCR abundance ranking from numerical data so that quantity of species can be indicated in mapping. As is the case for crude spatial information, the database should be able to filter out qualitative data. <i>MarLIN</i> provides a classification based upon the extent to which the data are quantitative or qualitative.</p>
<p>Biotope matching Automated biotope matching depends upon abundance scale data. <i>MarLIN</i> offers to take raw data and identify the closest biotope.</p>	<p>Work to develop a biotope matching programme is in progress by Dr Steve Wilkinson of JNCC. Biotope matching already exists for terrestrial ecosystems.</p>
<p>Keywords Various datasets are described in different ways.</p>	<p>Adopting existing standard dictionaries and thesauri allows more efficient keyword searches.</p>
<p>Temporal changes <i>MarLIN</i> is expected to identify species occurrence in the same location at different times.</p>	<p>So long as distribution information is georeferenced and its date of collection is recorded temporal change can be identified.</p>

European Register of Marine Species (ERMS) This project will produce a register of marine species in Europe, linked with other relevant information. Both the draft and final versions will be in the public domain and accessible through the World Wide Web (M. Costello, pers. Comm.).

The Marine Conservation Society (MCS)/Ulster Museum Species Directory for Britain and Ireland. At present, only the paper version is available and access to an electronic copy is essential if the system is to be applied as part of the NBN electronic dictionary. This code is probably the most comprehensive and widely used taxonomic checklist in academia in the U.K., however in its present form it is unsuitable for data manipulations within database applications.

The NODC (National Oceanographic Data Centre) Taxonomic Code, Version 8.0. This code (developed in the US) may not be comprehensive for Eastern Atlantic/ Mediterranean taxa, and is no longer supported by the National Oceanographic and Atmospheric Administration (NOAA).

UNICORN code. This is an internal code used by the UNICORN database. It is a numeric code that refers to a table containing species names.

6. Hardware/software issues

6.1 Software

Simultaneous interrogation of datasets over the World Wide Web will be enabled by software under development by JNCC. This software consists of Java objects, which are one of the technologies available for server-side processing (see section 4.4).

6.2 Location of the server

The server chosen for the purposes of the data access demonstration should be available to install the database engine and dll, and have the space to run its processes. To allay security concerns raised by the Environment Agency, and since doing so presents no technical problems, it may be best to. The advantages of this would be:

- the *MarLIN* server fulfils the required criteria;
- data providers can release exactly what level or section of data they want;
- convenience and resources are saved if *MarLIN* staff can work on the hardware and software locally;
- possible problems with the firewall at JNCC interfering with the operation of the database engine and dll could be avoided, and
- for the purposes of the data access demonstration (see section 4) remote access could still be demonstrated through JNCC dialling-in to the data at *MarLIN*.

7. Conclusions

Dynamic links to data will be achieved by mounting JNCC software on data providers servers. The Environment Agency would prefer to 'filter' the data prior to giving access to them, rather than have third party software running on their server to achieve this. This problem could be encountered with other data providers. Access over the Internet to the Environment Agency data is unnecessary given the size of the dataset and the frequency with which it is updated. A static copy of certain datasets on the *MarLIN* server maintained by periodic data transfer may be necessary. For the purposes of the data access demonstration the trial dataset should be held on the *MarLIN* server.

8. Acknowledgements

Dr Keith Hiscock guided the scoping study, Dr Lesley Rickards (BODC) and Dr James Williams (JNCC) provided advice on standards for cataloguing metadata. Dr Graeme Storey of the EA's NECDS at Bath provided his time and technical assistance, Dan Lear gave advice on hardware and software issues and Nigel Grist of Unicomarine gave technical information on the structure of UNICORN.

Appendix 1: Glossary of acronyms

BODC	British Oceanographic Data Centre.
CDS	Catalogue of Data Sources.
EA	Environment Agency.
EDMED	European Directory of Marine Environmental Data, developed by BODC.
EEA	European Environment Agency
FTP	File Transfer Protocol. A client/server protocol for exchanging files with a host computer.
GEMET	General European Multilingual Environment Thesaurus.
JNCC	Joint Nature Conservation Committee.
MNCR	Marine Nature Conservation Review, undertaken by JNCC.
NBN	The National Biodiversity Network.
NCEDS	The Environment Agency's National Centre for Environmental Data and Surveillance.
NMMP	National Marine Monitoring Programme.
UKDMAP	United Kingdom Digital Marine Atlas, developed by BODC.
UNICORN	Database developed by Nigel Grist of Unicomarine. It serves two main functions: Maintenance of taxonomic information and facilitation of data entry.
URL	Universal Resource Locator – an 'address' for an Internet website.

Appendix 2: Metadata Collection Form and Guidance notes.*MarLIN***Metadata collection form mandatory fields****Administrative Information**

1	Resource	Dataset/Document/Catalogue/Specimen or herbarium collection/Images/Other ...		
2	Originator			Acronym
3	Description			
4	Contact	Title	First name	Surname
5	Status	Unfinished /Complete / Ongoing / Other ...		
6	Compiler	Title	First name	Surname

Resource information

7	Language	
8	Title	
9	Abstract	

10	Purpose	
11	Date information	Next planned revision
12	Reference of report associated with resource	
13	Year published	

Availability

14	Access terms	Free/Licence/Purchase/Not available/Other...
15	Ordering information	
16	Cost	
17	Data support	
18	Copyright conditions	
19	Data format	
20	Size	
21	Storage OS	
22	Storage medium	
23	Storage quantity	

Organisation

24	Name		Acronym
25	Type		
26	Address 1		
27	Address 2		
28	Address 3		
29	County		
30	Post code		
31	Country		

32	Switchboard/ Fax	
33	URL	
34	Email	

Person Information

35	First name	
36	Last name	
37	Address 1	
38	Address 2	
39	Address 3	
40	County	
41	Post code	
42	Work phone	
43	URL	
44	E-mail	

Organisation

45	Name		Acronym
46	Position		

Keywords

47	Keywords		
48	MarLIN data classification	a) Type of data	b) Quality measure
49	Parameters		

Species

50	Checklist	
51	Source if non-standard checklist	
52	Taxa	

53	Person identifying	Title	First name	Last name

Biotopes

54	Classification			
55	Source			
56	Biotope	Code		

Temporal coverage

57	Start Date			
58	End Date			

Spatial coverage

59	Descriptive location			
60	County			
61	MNCR coastal sector			
62	Position of survey station			
	Bounding coordinates	Top left	Bottom right	

MarLIN**Metadata collection form mandatory fields - Guidance notes**

(Based on the National Biodiversity Network index and the European Directory of Marine Environmental Data)

Administrative Information**1. Resource type**

Dataset/Document/Catalogue/Specimen or herbarium collection/Images/Other

2. Originator

The organisation that owns the dataset or other form of resource (usually the organisation that paid for the collection of the data including supplying filmstock for images).

3. Description

A brief description of the role and activities of the data holding centre, and the data services it can provide. If the centre only releases or exchanges data by special arrangement this should be mentioned, as should any other special conditions, charges and procedures for the supply of data. Mention should also be made of any 'Users Guides' or publications available from the 'Contact Point' further describing the services and data holdings of the centre.

4. Contact

The person within the organisation responsible for the resource. This should be the person to whom enquiries are directed.

5. Status

Unfinished /Complete/ Ongoing/Other

6. Compiler

Name and title of individual.

Resource Information**7. Language****8. Title**

Title by which the resource (dataset) is formally known.

9. Summary (Entries should not exceed about 500 words)

The information covered will vary according to the data set but, where relevant, should at least include the following:

- a description of the nature of the measurements/samples, the purpose for which they were collected and the methods of sampling used in their collection;
- a statement on the level to which the data are processed and quality controlled and any known limitations on their reliability;
- The enumeration method used (e.g. abundance scales for conspicuous species, direct counts or measures in situ, counts of individuals in samples);
- the spatial and temporal resolution of data within the data set;
- information on the length of time series within the data set;
- estimate of the amount of data expressed in terms of the number of stations, sites, observations, or other units as appropriate;
- a statement of data sources i.e. which organisations contributed data to the data set. Are they all national or were foreign sources used to compile the data set?

10. Purpose

Why the data resource was created and what objectives it fulfils.

11. Date information

The date of last revision indicates when the resource was last updated. If the resource is updated regularly, place the date of the next revision in the 'next planned revision' field.

12. Reference

Include references to any information sheets or published literature that are available further describing the data set.

13. Year published**Availability****14. Access terms**

Free/Licence/Purchase/Not available/Other

15. Ordering information

How to order/obtain the resource or access to it. Also, any special order conditions.

16. Cost

If applicable, in pounds sterling.

17. Data support

Can range from 'None' to specific details.

18. Copyright conditions

Details of any such conditions or reference to their source elsewhere.

19. Data format

Spreadsheet, database, text file etc. Also please state program (Excel, Paradox, Unicorn etc.).

20. Size

Approximate Mb. If more than one file please indicate total.

21. Storage OS

DOS, Win .x, Win 95, NT, IBM, MAC, UNIX, other, N/A .

22. Storage Medium

Media, e.g. 3.5" disk, CD, Tape, Paper, Optical disk, Samples, Zip disk, Video, Record cards, other.

23. Storage quantity

Describe the quantity as the number of units of the medium (e.g. 10 magnetic tapes, 5 floppy disks, 6 optical disks, 16 printed volumes, 25 hard-copy maps, 70 microfiche reels, 600 boxes of analogue records, etc.). If the data set is spread over different media, entries should be made for each medium.

Organisation This information identifies organisations in the NBN index. If an email/URL is given then address info can be omitted and vice-versa.

24. Name

Full formal name of organisation and any official acronym.

25. Type

Statutory Nature Conservation Organisation, non-governmental, University, Commercial (public or private), Trust etc.

26. Address 1

Building name

27. Address 2

Street name

28. Address 3

Town name

29. County

30. Post code

31. Country

Defaults to UK if left blank

32. Switchboard/ Fax

'Phone numbers for the organisation. Note these should not be direct lines to a contact, whose number is provided below.

33. URL

If the organisation has a WWW site, provide the URL here

34. E-mail

If the organisation has a general E-mail account, provide details here. Note: should not be a personal E-mail to a contact, which is provided for below.

Person Information. The information is similar to that above for organisation, but is specific to a contact. If an organisation is provided and address information is given then it does not have to be re-presented here.

35. First name

36. Last name

37. Address 1

38. Address 2

39. Address 3

40. County

41. Post code

42. Work phone

43. URL

44. E-mail

Organisation. This indicates the role of the person within an organisation for which details have been given above.

45. Acronym and name

These are provided to indicate for which organisation the person works.

46. Position

The role of the person in the organisation

Keywords. Please list keywords that best describe the resource. If they are from a standard list then please give the source. In the future a standard list of keywords will be used to cover the resources of interest to the NBN. In the meantime two categories of keywords can be used. The first is a subset of the MNCR list of keywords (attached). The second is an uncontrolled list that can contain your own keywords.

47. Keywords

48. MarLIN data classification

Please a) Classify the dataset by selecting one of the seven types that best describes the way in which the data were collected,

- **Fully quantitative data** with all sampled taxa identified to distinctive entities including c. 80%+ to species. (This category usually applies only to samples collected from sediments where organisms occur as countable individuals. Where samples are sieved, 0.5 mm or 1 mm mesh will have been used and quality control measures taken.)

- **Semi-quantitative but comprehensive data** expressed as defined abundances or rough counts or % cover but for all conspicuous species. (All data which includes a mixture of solitary and colonial/crustose species.)
- **Semi-quantitative data from check-list of species.**
- **Semi-quantitative data from casual records.**
- **Qualitative but comprehensive data.** (Presence-absence data from surveys where all sampled taxa were identified to distinctive entities including c. 80%+ to species.)
- **Qualitative data from check list of species.**
- **Qualitative data from casual records.**

and b) Select a MarLIN quality measure from the following:

- **Good.** The data has been collected by a reputable institution with high accuracy standards. (This might be a laboratory which is part of the National AQC scheme)
- **Adequate.** The data is reliable but may include minor inaccuracies in species identification which would be easily spotted,
- **Limited value.** There are errors in species identification and or quantification which might not matter if species are anyway widely distributed.
- **Best guess / dubious.** There are most likely to be errors in species identification and or quantification which matter but the records are to be kept in case unusual records prove to be of historical importance.

49. Parameters

Checklist of the measured-variables/samples represented within the data set, using free-language key words that the holding centre finds appropriate.

Species. The NBN uses the standard NBN taxonomic dictionary. If the resource being described covers taxa please indicate which. Additionally please provide information about the checklist used to validate the species names, if known.

50. Checklist

Title of formal checklist used to define species name (i.e. BOU etc.)

51. Source

If a non-standard checklist is used please indicate the source.

52. Taxa

Taxa can be detailed at any level (i.e. Order, Family or species). Please give the latin binomial wherever possible. The common name is optional.

53. Person

Name of person responsible for identification, if possible.

Biotopes. The NBN uses the standard NBN biotope dictionary. If the resource being described covers biotopes and habitat types please indicate which. Additionally please provide information about the classification used.

54. Classification

Title of the formal classification used to define the biotope.

55. Source

If a non-standard classification is used please indicate the source.

56. Biotope

If possible provide the codes for all biotopes covered by the resource. The biotope name can be provided for cross checking.

Temporal Coverage. If the resource contains information related to time periods, the details can be provided here. If only one period is covered please indicate the start and end date for the whole resource.

57. Start date

58. End date

Spatial coverage.

59. Descriptive location

Commonly used but unambiguous place names, (e.g. Loch Long, Highland; Bull Bay, Anglesey; Salcombe Harbour; Dogger Bank).

60. County

61. MNCR coastal sector

Optional

62. Position of survey station / Bounding coordinates

Co-ordinates of survey stations can be entered as Ordnance Survey Grid References (to a maximum definition of 8 figures) or latitude and longitude (to a maximum definition of 0.001 minutes). If the resource contains information about a specific area that can be described by a bounding rectangle then please refer to this information here as co-ordinates for top left and bottom right corners.

Appendix 3 NMMP survey data, location/species count matrix, output from UNICORN as Microsoft Excel spreadsheet.

Appendix 4 UNICORN database relationships diagram