How can the NBN Atlas work better for marine data?

Notes from a workshop held at the NBN Conference in Cardiff on 16th November 2017
Summary produced by workshop facilitator Paula Lightfoot, p.lightfoot@btinternet.com

Workshop delegates:

- Charlotte Bolton, Seasearch
- Purba Choudhury, NBN Secretariat
- Katharine Davies – Greenspace Information for Greater London
- Moustafa Eweda, Cumbria Biodiversity Data Centre
- Guy Freeman, British Wildlife
- Natalie Harmsworth, The Wildlife Information Centre
- Manon Katell Jobic, Splatter Project
- Dan Lear, Marine Biological Association
- Paula Lightfoot, Seasearch
- Damian McFerran, Centre for Environmental Data and Recording
- Justine Millard, Marine Conservation Society
- Chris Raper, Natural History Museum
- Laura Sivess, Natural History Museum
- Fiona Ware, National Museums Scotland

Additional input was provided via e-mail by representatives of the Conchological Society of Great Britain and Ireland, Porcupine Marine Natural History Society, British Phycological Society, the NBN Secretariat and the Joint Nature Conservation Committee.

Agreed actions

1	Coordinate the addition of suitable marine base layers to the NBN Atlas	Dan Lear, MBA
2	Check that recent changes to the UKSI have been captured in WoRMs.	Dan Lear, MBA
3	Ensure that the Marine Recorder dictionary upgrade captures all amendments to the UKSI since the last upgrade.	Chris Raper, NHM JNCC
4	Discuss whether the UKSI can be updated on the NBN Atlas more frequently than at present, e.g. quarterly?	Chris Raper, NHM Sophia Ratcliffe, NBN Sec.
5	Ensure that proposals to introduce record commenting and verification status flags are suitable for marine data. Help to engage the marine community with this new functionality when it is available.	Paula Lightfoot Working Group 1
6	Coordinate a pilot project to enable experts to improve the accuracy of distribution data on the Atlas for a selection of marine species	Dan Lear, MBA Paula Lightfoot

Requests for development

1. Enable records from outside the British and Irish National Grids to be displayed on maps.
2. Make the 'marine' flag from the UK Species Inventory available as a filter for data visualisation and download.
3. Include more attributes in downloaded data or enable users to select attributes e.g. live/dead, specimen, photo.
4. Enable display of EUNIS habitat data as points and polygons.
5. Provide an interface for experts to query dubious records and ensure such queries remain appended to the record, i.e. are not overwritten when a dataset is updated.
6. Add locality and recorder name to the information that is shown in the preview table when records are selected. It is currently necessary to view records one at a time to get this information.
7. Make site boundary datasets discoverable under Data and Partners/Search NBN Atlas datasets.

NBN Atlas functionality

Taxon occurrence records from outside the British and Irish National Grids can be viewed as lists but cannot currently be displayed on maps. This is planned for future development. It is a priority for marine users as it will enable records from the UK Offshore Marine Area and the Channel Islands to be displayed.

The ability for users to upload their own site boundaries for data searches is useful, but participants felt that the interface could be made more user-friendly. Participants were unsure whether site boundaries for areas outside the British and Irish National Grid can be uploaded.

The NBN Atlas and GBIF use the Darwin Core Archive (DwC-A) data standard. DwC-A has the conceptual data model of a 'star schema' consisting of a single record at the centre of the star and optional extensions radiating from the central record (Figure 1).

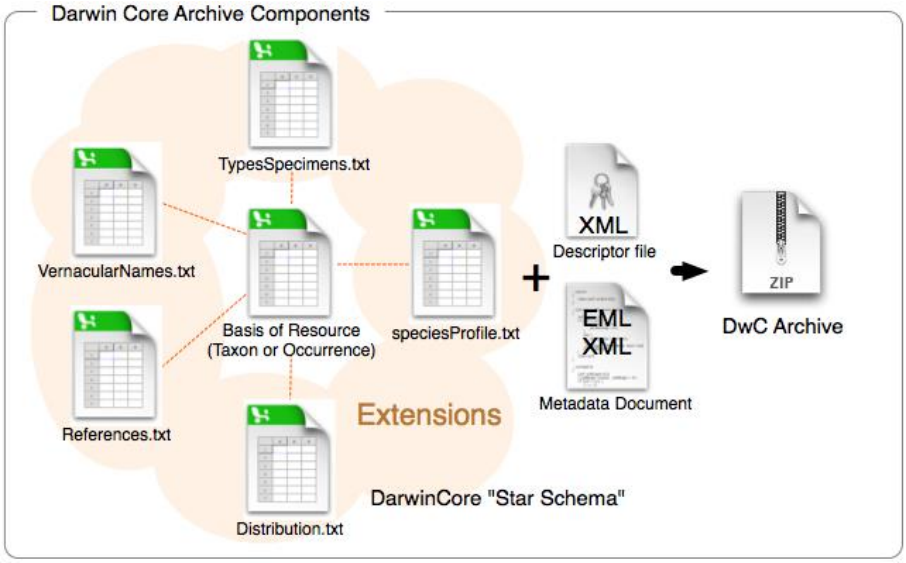

Figure 1: Darwin Core Archive 'star schema'. Source: https://gcube.wiki.gcube-system.org/gcube/Darwin_Core_Terms

This structure is not optimised for supporting rich marine data; a marine species occurrence record may be associated with sampling parameters, biotic and abiotic measurements, all organised in an event hierarchy (e.g. Survey > Event > Sample > Sub-sample). The Ocean Biogeographic Information System (OBIS) has built on the DwC-A standard to develop a data structure more suited to marine datasets which combine taxonomic data with biological, chemical and physical measurements (Figure 2).

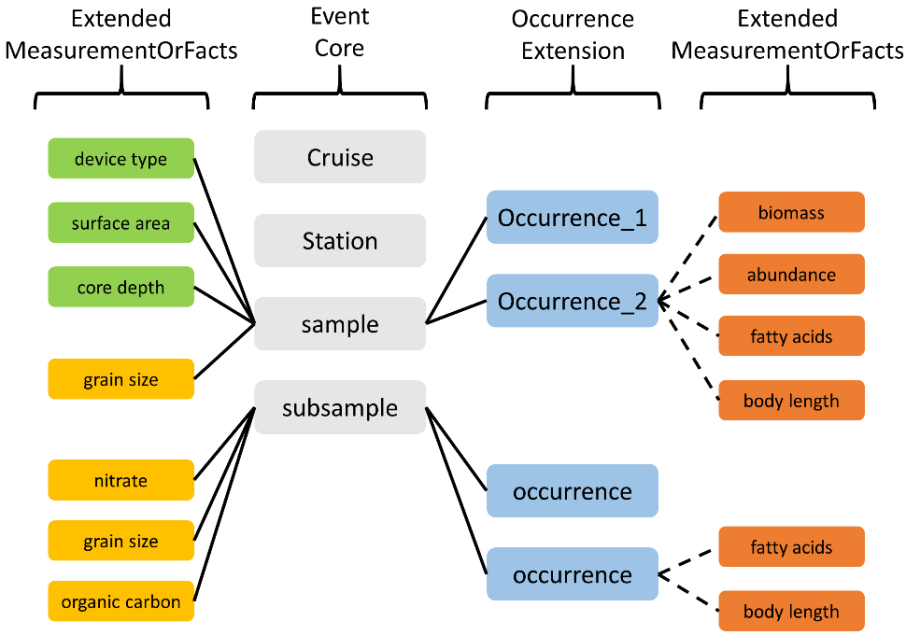

Figure 2: Overview of OBIS-ENV-DATA format. Source: http://www.iobis.org/manual/dataformat/

Participants discussed whether the NBN Atlas should adopt this enhanced data structure. It may not be a priority for development as marine data are a small proportion of total data on the Atlas and there are other portals for discovering and accessing marine data. On the other hand, the UK has statutory obligations for marine spatial planning, conservation, monitoring and reporting – the extent of protected areas at sea is over double the extent of protected areas on land. With new Atlases being developed, e.g. Northern Ireland, this discussion is timely and all options should be considered.

Participants discussed whether there should be a separate 'Marine Atlas'. This is technically possible because the Atlases are delivered by web services, and data in the main Atlas dataset could be filtered taxonomically and/or spatially. However, it would require financing and administration – it might be preferable to focus on enhancing the main Atlas for marine data.

The geographic coordinate system of the Atlas was discussed and clarified, because marine data tend to be georeferenced using latitude and longitude rather than Ordnance Survey grid references. Shapefile downloads from the Atlas are in WGS84 coordinate system. The attribute table contains a field 'raw_datum' which indicates whether a record was originally captured using a different system, e.g. OSGB. The field 'crdnt_ncrt' indicates the record's precision in metres.

Participants agreed it would be useful if the 'marine flag' in the UK Species Inventory could be used as a filter to display and download data. The 'marine flag' is already used in iRecord as a filter for verifiers who wish to verify only marine or non-marine taxa within a larger group, e.g. Crustacea or Mollusca. It is not perfect and users are requested to let Chris Raper know if they come across any species incorrectly flagged as 'marine' or vice versa.

It would also be useful if records could be filtered by the attribute 'live/dead', this is particularly relevant for mollusc records which may be recorded from empty shells.

Contextual marine data

The base layers currently on the NBN Atlas do not provide sufficient context for marine data. Licensing costs prevent use of UK Hydrographic Office (UKHO) Admiralty Charts. However, useful base maps are available free or at reduced cost that could be added to the Atlas or delivered via Web Map Services (WMS) e.g.

- OceanWise marine and coastal mapping data: https://maps.oceanwise.eu/
- General Bathymetric Chart of the Oceans (GEBCO): https://www.gebco.net/ and https://www.gebco.net/data_and_products/gebco_web_services/
- EUSeaMap benthic habitat maps available via WMS from European Marine Observation and Data Network (EMODnet): http://www.emodnet-seabedhabitats.eu

Marine Scotland's National Marine Plan Interactive website uses OceanWise Marine Themes DEM and Raster Charts via WMS as base layers: https://marinescotland.atkinsgeospatial.com/nmpi/

Action 1	Coordinate the addition of suitable marine base layers to the NBN Atlas	Dan Lear, MBA

Delivering habitat maps via WMS would not enable data to be queried e.g. 'what species occur on a particular habitat?' or downloaded. It might therefore be preferable to deliver habitat maps as layers or via Web Feature Services (WFS). The Atlas currently has no marine habitat data. Mobilisation of terrestrial habitat data is challenging because there are numerous terrestrial habitat classification systems, however, the Marine Habitat Classification for Britain and Ireland (v15.03) is compatible with and contributes to the European Nature Information System (EUNIS) Habitat Classification. If the functionality could be developed to display habitat data using these systems (or just EUNIS), marine habitat data could be mobilised, including:

- Point data from the Marine Recorder database
- Polygon data from EMODnet
- Polygon data from Natural England surveys via WFS from data.gov.uk

A wide range of marine spatial environmental data is freely available as downloadable layers or via WFS from various providers and archives. A few examples are:

- https://www.thecrownestate.co.uk/rural-and-coastal/coastal/downloads/maps-and-gis-data/
- https://odims.ospar.org/
- https://www.cefas.co.uk/cefas-data-hub/
- https://www.channelcoast.org/

The English Marine Conservation Zone layer on the Atlas contains a mixture of designated and recommended sites and it is not possible to filter them on designation status. Site boundaries of potential Marine Protected Areas (MPAs) are extremely useful and help to focus survey effort, but we need to be able separate designated sites from recommended sites. Site boundaries on the Atlas must be consistent with those available on the JNCC MPA interactive map: http://jncc.defra.gov.uk/page-5201

Marine taxonomy and nomenclature

The following points about marine taxonomy and nomenclature were discussed and clarified:

- The World Register of Marine Species (WoRMS) is a global database of names of marine taxa. It consists of separate taxonomic checklists maintained by the relevant experts. It is the official taxonomic reference list for OBIS and a component of GBIF Backbone Taxonomy. http://www.marinespecies.org/
- Marine Species of the British Isles and Adjacent Seas (MSBIAS) is a geographic subset of the WoRMS database. http://www.marinespecies.org/msbias/
- The UK Species Inventory (UKSI) is a database of names of UK taxa from all habitats. It is managed by Chris Raper at the Natural History Museum. Individual checklists are supplied by the relevant experts. It is the taxonomic backbone of NBN Atlas and iRecord. http://www.nhm.ac.uk/our-science/data/uk-species.html
- To data transfer between systems, each taxon has a unique identifier. These are TaxonVersionKeys (TVKs) in the UKSI and Aphia IDs in WoRMS.

The NHM, JNCC and MBA collaborated to integrate and align MS BIAS and the UKSI a few years ago. When new marine species names are added as a result of new records or taxonomic revision, both the UKSI and the master copy of WoRMS/MSBIAS must be updated. It is possible that some recent updates to the UKSI have not been captured in WoRMS/MSBIAS, but this can be rectified.

JNCC have advised that the Species Dictionary in Marine Recorder is currently being updated, aiming for release by the end of 2017. We need to ensure that all recent changes to the UKSI are captured in this update, to ensure that records of 'new' species can be shared via the Atlas without delay.

The UKSI used to be updated on the Gateway roughly every 6 months and this is continuing on the Atlas. It was agreed it would be worth investigating if this could be done more frequently. It was thought unlikely that these updates could be automated.

Taxonomic revisions need to be communicated to dataset managers so that they can take appropriate action with records in their database if necessary, e.g. assign records to a *sensu lato* aggregate. Can this communication be automated, e.g. dataset administrators could sign up to receive notifications of changes to species with the 'marine' flag?

Action 2	Check that all recent changes to the UKSI have been captured in WoRMs.	Dan Lear, MBA
Action 3	Ensure that the Marine Recorder dictionary upgrade captures all amendments to the UKSI since the last upgrade.	Chris Raper, NHM JNCC
Action 4	Discuss whether the UKSI can be updated on the NBN Atlas more frequently than at present, e.g. quarterly?	Chris Raper, NHM Sophia Ratcliffe, NBN Secretariat

Marine data quality

The Atlas contains erroneous taxon occurrence records in many marine datasets. These have arisen as a result of:

- Misidentification
- Coordinate data entry error e.g. marine records on land
- Absence accidentally entered as presence
- Selection of wrong species name from drop-down list
- Taxonomic revision

It was agreed that it would be desirable to 'clean up' marine data; two approaches were discussed:

1. Atlas interface for verification

   The Gateway had a 'record commenting' facility enabling any logged-in user to query a record and append comments explaining the reasons. This sent an email to the dataset administrator and the query was visible to all users of the Gateway. However, the query would be lost if the dataset was updated as the new version of the dataset would simply overwrite the old one.

   The ability to query records on the Atlas would be welcome, but it is important that queried records are not simply overwritten if a dataset is updated.

   The system would need the flexibility to distinguish between records which are simply wrong and need to be deleted or changed, and those which are questionable and cannot be confirmed but should not be removed in case they are correct.

   The ability to comment could be restricted to appointed experts, e.g. using the structure already in place on iRecord. However, the marine recording community isn't organised into taxonomic recording schemes and county recorders so appointing experts might not be straightforward. Furthermore, you don't need to be an 'expert' to spot some errors. It might therefore be worth enabling any registered users to comment on records, provided that this isn't abused.

   NBN Strategic Working Group 1 is already working on proposals to introduce record commenting and machine readable verification status flags.

Action 5	Ensure that proposals to introduce record commenting and verification status flags are suitable for marine data. Help to engage the marine community with this new functionality when it is available.	Paula Lightfoot Working Group 1

2. Pilot project to improve the accuracy of distribution data for a selection of marine species:

i. Establish a working group of experts to deliver the project
ii. Identify a manageable list of species to tackle. These will be selected based on concerns about the accuracy of their distribution data and the importance of knowing their distribution, e.g. because they are climate change indicators, non-natives or conservation priorities. It could also include species whose distribution is uncertain due to recent taxonomic revision.
iii. Establish a workflow for evaluating the distribution data currently available via the Atlas for each species and documenting the output of this evaluation.
iv. Contact dataset administrators regarding records queried as dubious and requiring confirmation. This could be done via the Atlas if new record commenting functionality is developed, or offline if not. The NBN Secretariat have agreed to take action on queried records if there is no response from dataset administrators after an agreed period, e.g. remove from default display.
v. Document outcomes i.e. records removed or changed.
vi. Communicate results to the wider NBN community.
vii. Depending on the success of the project, perhaps repeat with a new list of species.

   Although this would only improve data quality for a small number of species, it would raise awareness that we (the community) are responsible for data quality, quantify the scale of the

problem and the effort needed to tackle it, and produce a documented workflow that others could adopt. It could also provide the impetus to engage people in improving data quality on an ongoing basis via record commenting once the pilot project is over.

The MBA are willing to coordinate this effort, and representatives from several marine recording schemes have already indicated they are willing to get involved. Paula Lightfoot will help as required.

Action 6	Coordinate a pilot project to enable experts to improve the accuracy of distribution data on the Atlas for a selection of marine species	Dan Lear, MBA Paula Lightfoot

Record Cleaner rules already exist for a number of marine species but several key groups are not represented (e.g. Mollusca) and the distribution rules are out-of-date for many species. Expanding the taxonomic coverage of the rulesets could help to reduce the number of new erroneous records. The functionality to update distribution rules automatically when new verified records extend a species' known distribution would be desirable.

Marine Recorder includes an attribute field to state whether a voucher specimen or photograph exists for the record. This field does not appear to be available as a filter on the Atlas or in downloaded data.

Marine data flow

The following aspects of marine data flow were clarified:

- Data from the Atlas are exported to GBIF; the last export was done just prior to the conference.
- Marine data are not exported from the Atlas to OBIS, but are exported from the Archive for Marine Species and Habitats Data (DASSH) to OBIS. Data on OBIS are less up-to-date than data on the Atlas, e.g. the Seasearch dataset on OBIS includes records only up to 2014.
- Marine Recorder used to export data in NBN Exchange Format for sharing to the NBN Gateway. The MBA have developed a tool to export from Marine Recorder in DwC-A format for the Atlas. JNCC have stated that they are moving Marine Recorder to an online open source platform in 2018.
- Records of marine species entered into iRecord directly or via other Indicia websites or apps are verified by MBA staff or external experts and shared via the Atlas. Most are in a dataset administered by the MBA entitled Verified marine records from Indicia-based surveys. Marine data collected for a particular project or recording scheme may be shared in separate datasets administered by the project or scheme organiser.

Marine data mobilisation

There was a brief discussion about marine datasets that are not currently available via the Atlas. Some would be difficult to mobilise due to issues of data protectionism, others may become accessible (e.g. via the reciprocal agreement between DASSH and the NBN) after an embargo period, e.g. data from commercial surveys.

Some examples include:

- Cetacean data from ferry surveys
- Seahorses
- Seabirds
- Plankton time-series
- Data from Environmental Impact Assessments e.g. for offshore windfarms

Many thanks to everyone who participated in the workshop and who sent in suggestions and comments via e-mail!