EMODnet Seabed Habitats

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241

Start date of the project: 25/09/2019 –(24 months)

EMODnet Phase III extension

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

Disclaimer

Document info

Title	A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension
WP title [ref]	WP2 – Data Collation [WP2]
Task [ref]	Guidance for submitting habitat point data to OBIS [D2.08]
Authors [affiliation]	Graeme Duncan [JNCC], Dan Lear [MBA], Kevin Paxman [MBA], Helen Lillis [JNCC], Lewis Castle [JNCC]
Dissemination level	Public
Deliverable due date (if applicable)	15/07/2021
Keywords and/or short description	Technical guidance on using the Darwin Core Extended Measurement or Fact Extension in the OBIS-ENV-DATA approach to submit categorised habitat data into EurOBIS maintaining links with associated species occurrences.

Document history

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

(Version)	Authors [affiliation]	Date
1.0	Graeme Duncan [JNCC], Dan Lear [MBA], Kevin Paxman [MBA], Lewis Castle [JNCC]	30/06/2021
1.1	Helen Lillis [JNCC]	10/07/2021
Final [1.2]	Graeme Duncan [JNCC]	25/07/2021

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the
Darwin Core Extended Measurement or Fact Extension

Contents

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the
Darwin Core Extended Measurement or Fact Extension

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

1 Introduction

With aims of facilitating international interoperability between disparate benthic habitat data holdings, and improving interoperability and standardisation in increasingly common holistic benthic surveys, capturing both species and habitat information (among other data types), EMODnet Seabed Habitats has worked alongside EMODnet Biology and Ocean Biodiversity Information System[1] (OBIS) colleagues to produce guidance for submitting habitat data in listed classification systems into the OBIS ecosystem.

Recent developments in the Darwin Core[2] international biodiversity data standard include both a move towards an optional event-driven data structure ("Event Core") and the ability to record additional measurements or facts against both events and species occurrences.

Both developments provide the framework for the low-friction submission of habitat data in the EurOBIS[3] European node of OBIS, alongside species occurrence information, bringing together these two important data types and enabling holistic biodiversity datasets to be aggregated at a European scale.

A standard approach to structuring habitat measurements is key to making it discoverable, understood and reusable. This short guidance document describes a recommended approach agreed between EMODnet Seabed Habitats, EMODnet Biology and OBIS for structuring habitat information in preparation for submission to EurOBIS.

Within this document, the term "habitat" is used in the broadest sense, covering physical habitats, biotopes, biocenoses and communities.

2 The OBIS ENV-DATA approach – a brief overview

With a move to capturing valuable additional data and parameters alongside species occurrences, EurOBIS now accepts data in the "Event Core" data structure. In this structure, the central unit around which other records relate is an "event". An event can be at any level of granularity (e.g., cruise, transect, sample, subsample), and events can be hierarchical, linking up to parent events in the case of a multi-sample structured dataset. Occurrences are then associated with the most granular event in which they occurred.

---

[1] https://obis.org/

[2] https://dwc.tdwg.org/

[3] https://www.eurobis.org/

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the
Darwin Core Extended Measurement or Fact Extension

With this core approach, the "ExtendedMeasurementOrFact"[4] (eMoF) extension, can then be used to associate any relevant measurement or fact against either an Occurrence or Event (see **Error! Reference source not found.**), so long as the measurement or fact is adequately and consistently documented and described, ideally associated with a managed and publicly accessible controlled vocabulary such as the NERC Vocab Server.
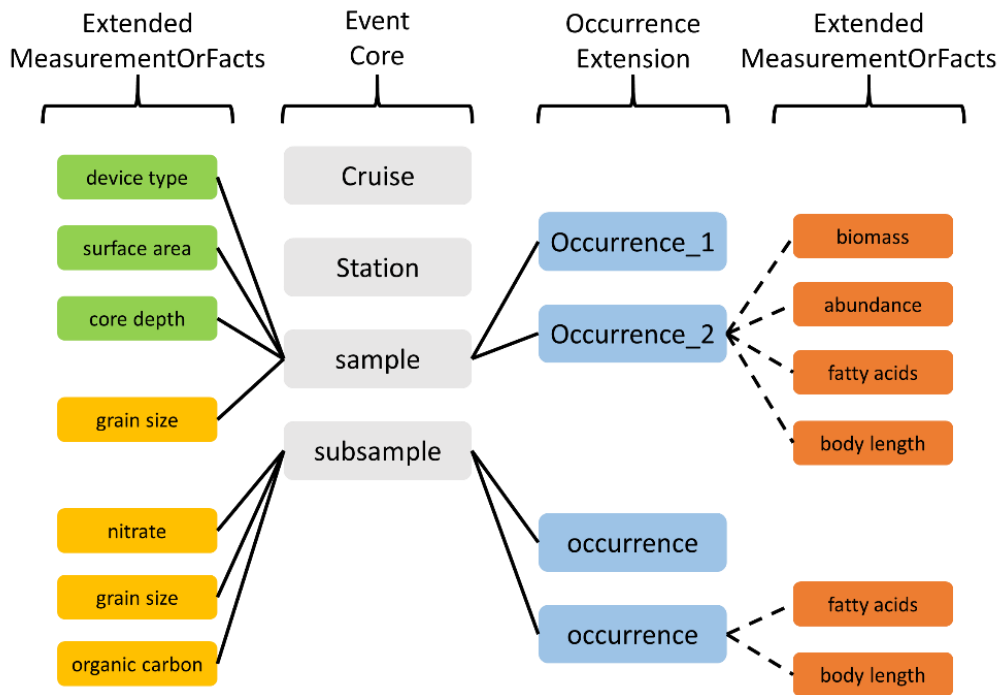

Figure 1. Overview of the ENV-DATA approach from https://obis.org/manual/dataformat

Further detailed information on the ENV-DATA approach can be found at https://obis.org/manual/dataformat.

3 Submitting categorised habitat data using Event Core and eMoF

Submitting categorised habitat information using the ENV-DATA is no different at the abstract level than other measurements of biodiversity.

---

[4] An adaptation of the original "MeasurementOrFact" extension, which could, among other differences, only be associated with Occurrences.

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

3.1 Structuring events

When providing habitat data under EMOF, events should be presented in enough granularity to provide both precision of the location and time of the habitat occurrence, and to allow for multiple habitat occurrences to be determined.

Most habitats are recorded at the sample or subsample event level, and as such a valid event record should be created for every habitat observation, with the corresponding sample or subsample information attached and sufficient detail to allow and end-user to accurately determine the habitat presence. Each eMoF record of a habitat then relates to its specific event only.

For example, if different habitats occurred along segments of a video tow, an event should be created for each segment, to allow maximum granularity, with the corresponding habitat eMoF record relating to the segment event identifier.

Including parent events is preferred and encouraged, for good data stewardship, but is not essential.

3.2 Structuring habitats measurements

The structure of a generic record in the eMoF table is detailed in Table 1.

Table 1: A generic description of each field contained within eMoF table.

Field	Description
id	The Event Id under which the measurement is associated
measurementID	An identifier for the MeasurementOrFact (information pertaining to measurements, facts, characteristics, or assertions). May be a global unique identifier or an identifier specific to the data set.
occurrenceID	The identifier of the occurrence the measurement or fact refers to. If not applicable, it should be left empty.
measurementType	The nature of the measurement, fact, characteristic, or assertion. Recommended best practice is to use a controlled vocabulary.
measurementTypeID	An identifier for the measurementType (global unique identifier, URI). The identifier should reference the measurementType in a vocabulary. It is recommended to use a term from the "P01" vocabulary on the NERC vocabulary server where possible.
measurementValue	The value of the measurement, fact, characteristic, or assertion.
measurementValueID	An identifier for facts stored in the column measurementValue (global unique identifier, URI). This identifier can reference a controlled vocabulary (e.g. for sampling instrument names, methodologies, life stages) or reference a methodology paper with a DOI. When the measurementValue refers to a value and not to a fact, the measurementvalueID has no meaning and should remain empty.

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the
Darwin Core Extended Measurement or Fact Extension

measurementUnit	The units associated with the measurementValue. Recommended best practice is to use the International System of Units (SI).
measurementUnitID	An identifier for the measurementUnit (global unique identifier, URI). The identifier should reference the measurementUnit in a vocabulary.
measurementAccuracy	The description of the potential error associated with the measurementValue.
measurementDeterminedDate	The date on which the MeasurementOrFact was made. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E)
measurementDeterminedBy	A list (concatenated and separated) of names of people, groups, or organizations who determined the value of the MeasurementOrFact.
measurementMethod	A description of or reference to (publication, URI) the method or protocol used to determine the measurement, fact, characteristic, or assertion.
measurementRemarks	Comments or notes accompanying the MeasurementOrFact.

Within the context of the description of habitats within defined classifications systems, the general approach taken is that:

- the measurementType asserts that **the measurement is of a habitat within the context of a specific classification system** itself
- the **measurementValue contains the class of the habitat** being described.

The use of mesurementTypeID and measurementValueID then remain as described in the eMoF standard, providing an identifier, preferably machine readable, to the term. A measurement describing a habitat is also unitless.

Therefore, when providing measurement records in the context of habitats, the eMoF fields should be used as detailed in Table 2.

Table 2: A description of each field contained within eMoF table, in the context of describing habitat data.

Field	Description
id	The Event Id in which the habitat occurred or was observed.
measurementID	A unique identifier for the measurement
measurementType	A text string describing that the measurement is a description of habitat by classification to a term, and including both the classification system name and classification system version used.

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

	For example: *"Description of habitat by classification to a term from the EUNIS habitat classification system (version 2007)"*
measurementTypeID	A machine-readable URI or DOI reference describing the (version of the) classification system itself. *For example:* *"https://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/"*
measurementValue	The class of habitat being described. If the habitat classification makes use of short-codes for its terms (such as EUNIS), both the code and class name should be used, to aid the end-user in understanding the value. *For example:* *"A5.36 (Circalittoral fine mud)"*
measurementValueID	If available, a machine-readable URI describing the habitat class in "measurementValue". *For example:* *"https://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/A5.36"*
measurementAccuracy	Remarks concerning the accuracy of the determination of the habitat class, such as may affect confidence in the habitat. *For example:* *"Uncertain – Biological zone inferred from modelled data"*
measurementDeterminedDate	The date on which the habitat class was attributed to the observation. Note this is not necessarily identical to the date of sampling, for example in videography that is analysed post-cruise.
measurementDeterminedBy	The individual(s) or organisation(s) who attributed this habitat class to the observation. Note that this field may contain personal data, it is up to the organisation submitting data to EurOBIS to process these personal data correctly.
measurementMethod	A description of the method used to determine the habitat class, if an identifiable method was undertaken.

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

	For example: *"Cluster analysis; Expert judgement"*
measurementRemarks	Any further remarks on the habitat or its determination.

3.3 Habitats described under multiple classification systems

Any habitat occurrence may be, and often is, described under more than one classification system, where such classifications cover different aspects of the habitat, are used for specific reporting requirements, or cover a different geographic scope (e.g. National and European classification systems).

This situation is adequately handled under ENV-DATA by creating separate eMoF records for each habitat description, while referring to the same id for the relevant event. If habitats were directly translated between classification systems (for example using a lookup table), this should be recorded in the measurementMethod field of the derived habitat, noting the measurementID of the source habitat if possible.

3.4 Example eMoF table

An example of how the eMoF table can be populated in the context of submitting habitat data is given in Table 3. The example provides insight into how records can be split by transect ("VidTR11" and "VidTR38" in the example), video segments ("SegA" and "SegB" for each transect in the example) and habitat type.

Within these samples, the habitat was first categorised under the Marine Habitat Classification for Britain and Ireland, before being translated into EUNIS using a lookup table. A few years later, these records were determined to be examples of biogenic Reef under the EU's Habitats Directive and each subsequently categorised for a third time.

3.5 Recommended values for submission of data under common classification systems

Last updated: 2021-06-21.

A list of recommended values for users to submit under the 'measurementType', 'measurementTypeID' and 'measurementValueID terms list' fields are detailed in Table 4. For identifiers that currently do not have a machine-readable endpoint, EMODnet Seabed Habitats are working with existing infrastructures to generate terms, and will update this table as new codelists and terms become available.

Table 3. Example of how the eMoF table is populated in the context of submitting habitat data.

id	measurementID	measurementType	measurementTypeID	measurementValue	measurementValueID	measurementAccuracy	measurementDeterminedDate	measurementDeterminedBy	measurementMethod	measurementRemarks
Surv1_VidTR11_SegA	Surv1_meas612	Description of habitat by classification to a term from EUNIS (version 2007-11)	https://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/	A5.611 (Sabellaria spinulosa on stable circalittoral mixed sediment)	http://dd.eionet.europa.eu/vocabularyconcept/biodiversity/eunishabitats/A5.611	Certain	01/02/2010	JNCC, UK	Derived via lookup from UK habitat class (Surv1_meas613) - exact match	
Surv1_VidTR11_SegA	Surv1_meas613	Description of habitat by classification to a term from the JNCC's Marine Habitat Classification for Britain and Ireland Version 15.03 (SDN:M21::)	http://vocab.nerc.ac.uk/collection/P01/current/UKMH1503/	SS.SBR.PoR.SspiMx (Sabellaria spinulosa on stable circalittoral mixed sediment)	http://vocab.nerc.ac.uk/collection/M21/current/JNCCMNCR00001112/	Certain	01/02/2010	JNCC, UK	Expert judgement	
Surv1_VidTR11_SegA	Surv1_meas614	Description of habitat by classification to a term from EU Habitats Directive Annex I	http://dd.eionet.europa.eu/vocabulary/biodiversity/n2000habitats/	1170 (Reefs)	http://dd.eionet.europa.eu/vocabularyconcept/biodiversity/n2000habitats/1170	No reefiness assessment undertaken	11/12/2018	JNCC, UK	Derived via lookup from UK habitat class (Surv1_meas613) - source class contained within destination class	Biogenic reefs form a subtype of Annex 1 reef, safe translation between habitats.
Surv1_VidTR11_SegB	Surv1_meas615	Description of habitat by classification to a term from EUNIS (version 2007-11)	https://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/	A5.625 (Mytilus edulis beds on sublittoral sediment)	http://dd.eionet.europa.eu/vocabularyconcept/biodiversity/eunishabitats/A5.625	Certain	01/02/2010	JNCC, UK	Derived via lookup from UK habitat class (Surv1_meas616) - exact match	

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

id	measurementID	measurementType	measurementTypeID	measurementValue	measurementValueID	measurementAccuracy	measurementDeterminedDate	measurementDeterminedBy	measurementMethod	measurementRemarks
Surv1_VidTR11_SegB	Surv1_meas616	Description of habitat by classification to a term from the JNCC's Marine Habitat Classification for Britain and Ireland Version 15.03 (SDN:M21::)	http://vocab.nerc.ac.uk/collection/P01/current/UKMH1503/	SS.SBR.SMus.MytSS (Mytilus edulis beds on sublittoral sediment)	http://vocab.nerc.ac.uk/collection/M21/current/JNCCMNCR00001202/	Certain	01/02/2010	JNCC, UK	Expert judgement	
Surv1_VidTR11_SegB	Surv1_meas617	Description of habitat by classification to a term from EU Habitats Directive Annex I	http://dd.eionet.europa.eu/vocabulary/biodiversity/n2000habitats/	1170 (Reefs)	http://dd.eionet.europa.eu/vocabularyconcept/biodiversity/n2000habitats/1170	No reefiness assessment undertaken	11/12/2018	JNCC, UK	Derived via lookup from UK habitat class (Surv1_meas616) - source class contained within destination class	Biogenic reefs form a subtype of Annex 1 reef, safe translation between habitats.
Surv1_VidTR38_SegB	Surv1_meas618	Description of habitat by classification to a term from EUNIS (version 2007-11)	https://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/	A5.611 (Sabellaria spinulosa on stable circalittoral mixed sediment)	http://dd.eionet.europa.eu/vocabularyconcept/biodiversity/eunishabitats/A5.611	Certain	21/06/2011	JNCC, UK	Derived via lookup from UK habitat class (Surv1_meas619) - exact match	
Surv1_VidTR38_SegB	Surv1_meas619	Description of habitat by classification to a term from the JNCC's Marine Habitat Classification for Britain and Ireland Version 15.03 (SDN:M21::)	http://vocab.nerc.ac.uk/collection/P01/current/UKMH1503/	SS.SBR.PoR.SspiMx (Sabellaria spinulosa on stable circalittoral mixed sediment)	http://vocab.nerc.ac.uk/collection/M21/current/JNCCMNCR00001112/	Certain	21/06/2011	JNCC, UK	Expert judgement	

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

id	measurementID	measurementType	measurementTypeID	measurementValue	measurementValueID	measurementAccuracy	measurementDeterminedDate	measurementDeterminedBy	measurementMethod	measurementRemarks
Surv1_VidTR38_SegB	Surv1_meas620	Description of habitat by classification to a term from EU Habitats Directive Annex I	http://dd.eionet.europa.eu/vocabulary/biodiversity/n2000habitats/	1170 (Reefs)	http://dd.eionet.europa.eu/vocabularyconcept/biodiversity/n2000habitats/1170	No reefiness assessment undertaken	11/12/2018	JNCC, UK	Derived via lookup from UK habitat class (Surv1_meas619) - source class contained within destination class	Biogenic reefs form a subtype of Annex 1 reef, safe translation between habitats.

Table 4: List of recommended values when submitting habitat data under common classification systems.

measurementType	meaurementTypeID	measurementValueID terms list	Comment
EUNIS (version 2007-11)	https://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/	http://dd.eionet.europa.eu/vocabulary/biodiversity/eunishabitats/view	measurementTypeID not currently machine readable
Habitats Directive Annex I Habitats	http://dd.eionet.europa.eu/vocabulary/biodiversity/n2000habitats/	http://dd.eionet.europa.eu/vocabulary/biodiversity/n2000habitats/view	measurementTypeID not currently machine readable
Marine Strategy Framework Directive Benthic Broad Habitat Types	http://vocab.nerc.ac.uk/collection/P01/current/MSPDHAB1/	http://dd.eionet.europa.eu/vocabulary/msfd/broadHabitatTypes/view	machine-readable vocabulary
HELCOM lists of threatened and/or declining species and biotopes/habitats in the Baltic Sea area	http://helcom.fi/Lists/Publications/BSEP113.pdf	N/A	PDF only
HELCOM Underwater biotope and habitat classification system (HELCOM HUB)	http://vocab.nerc.ac.uk/collection/P01/current/HELCHUB1/	http://vocab.nerc.ac.uk/collection/M23/current/accepted/	machine-readable vocabulary
OSPAR List of Threatened and/or Declining Species and Habitats	http://vocab.nerc.ac.uk/collection/P01/current/OSPARTD1/	http://vocab.nerc.ac.uk/collection/M22/current/accepted/	machine-readable vocabulary
Barcelona Convention - Classification of Benthic Marine Habitat Types for the Mediterranean Region	https://wedocs.unep.org/bitstream/handle/20.500.11822/7081/99ig12_5_eng.pdf	N/A	PDF only
Folk (1954) sediment classification	https://doi.org/10.1086/626171	N/A	DOI only
Marine Habitat Classification for Britain and Ireland v15.03	http://vocab.nerc.ac.uk/collection/P01/current/UKMH1503/	http://vocab.nerc.ac.uk/collection/M21/current/accepted/	machine-readable vocabulary

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241 – *EMODnet Seabed Habitats*

A standard approach to structuring classified habitat data using the Darwin Core Extended Measurement or Fact Extension

| **Peres Picard 1964** | http://paleopolis.rediris.es/BrachNet/REF/Download/Manuel.html | N/A | PDF only |

4 Submitting data to EurOBIS

The preferred method for sharing Darwin Core datasets to EurOBIS utilizes the Integrated Publishing Toolkit (IPT) application, developed by the Global Biodiversity Information Facility (GBIF). This free tool allows Darwin Core Archive (DwC-A) datasets to be shared publicly and additionally performs basic checks on the integrity of the dataset and provides a version control system. Additionally, the IPT supports the capability to allocate Digital Object Identifiers (DOIs) to the data, facilitating citation and reuse. The IPT is a free, open-source software Java application that any organisation can install.

However, installation and initial set-up is non-trivial, as such, for organisations and data providers that do not manage their own IPT installation, it is recommended to utilize another organisation's existing IPT installation. OBIS maintains a network of nodes with existing IPT installations, including a set of national and thematic nodes, which we recommend as an initial entry point. A list of regional and national OBIS nodes can be found at https://obis.org/contact/. Alternatively, if you are uncertain of the correct route or means of submission, we advise contacting EMODnet Ingestion (https://www.emodnet-ingestion.eu/questions) or using their data submission service (https://www.emodnet-ingestion.eu/data-submission) to guide you along the correct process.

4.1 EurOBIS publication and harvest process

EurOBIS perform harvests of new datasets roughly four times a year. These dates are not currently published in advance, so to ensure timely submission the EurOBIS Data Management Team should be contacted to find the date of the next harvest. For a dataset to be definitively included in a given harvest, EurOBIS have requested that the finished dataset is submitted at least two weeks prior to the harvest date. In addition, the dataset must pass their online validator tool (https://rshiny.lifewatch.be/BioCheck) without major errors. The BioCheck tool must be pointed at the URL of a published IPT dataset in order to function. As there is not as yet a comprehensive list of the requirements, and the source of the issues are often unique, creating a dataset that passes the validator tool usually requires an iterative approach with feedback both from the tool and from the EurOBIS Data Management Team. Datasets should therefore aim to be initially published a significant amount of time before the deadline for validated datasets, which, as noted, is in turn two weeks before the published harvest date itself.

Datasets can be uploaded to an IPT in two ways, either as a complete DwC-A, which is a zipped file containing CSVs of the data tables described above (event, occurrence, extendedMeasurementorFact), and the meta.xml and eml.xml files (which describes the structure of the DwC-A and provide the dataset's metadata respectively). Alternatively, just the data tables can be uploaded, and these XML files will be generated by the IPT automatically once the user has filled in the metadata fields through the tool. There is a minimum of mandatory metadata that needs to be added to a published dataset, including Title, Abstract, Organisations, Data Licence and Citation. A DOI is also required by EurOBIS - this can be added by the IPT if linked to your organisation's DOI account. If your organization is not publishing to an IPT directly, this metadata will need to be provided to the relevant national OBIS node.

Subsequent to the harvest date, there will be an additional number of weeks as the dataset is integrated into the current EurOBIS holdings. Additional checks will be performed at this point, and EurOBIS have stated it is still possible for a dataset to be rejected this stage due to earlier undetected issues. After this period, the dataset will be available through the OBIS portal. In addition, the IPT allows simultaneous publication on GBIF if the relevant option is selected. This supports the proposed "Publish Once, Harvest Many Times" paradigm, promoted through the agreement between OBIS and GBIF in support of closer collaborative working.