Joint Cetacean Data Programme

Phase One - Final Report

Version 1.4

Acknowledgments

The project was managed by JNCC and thanks its partner organisations Bangor University and Sea Watch Foundation.

This project was funded by the Defra R&D Healthy and Biologically Diverse Seas Evidence Group (HBSDEG) funding call for 2019/2020.

We thank the Steering Group participants for their input and review, as well as wider stakeholders who have contributed to the development of this project and associated outputs to date.

Executive Summary

Currently, there is no mechanism to readily access the wide range of existing cetacean monitoring datasets in the UK and wider northeast Atlantic waters, and the process to do so can be labour intensive and inefficient. Comprehensive data on which to perform analyses regarding the status of cetaceans is not available from a single source, but instead, agreement to use existing data needs to be brokered with multiple data owners/organisations. The Joint Cetacean Data Programme (JCDP) aims to collate these data into a single resource to maximise effectiveness of the various data collection programmes and vastly improve access to the cetacean evidence base. The JCDP is an effective method to turn short-term investment in standardisation and collation of data, into a long-term resource that will enable more robust analyses to inform policy, research and conservation. The resource would be accessible by all stakeholders, including NGOs, industry, research organisations and government bodies.

Phase one of the JCDP (2019 – 2020), funded by DEFRA through the Healthy and Biologically Diverse Seabed Evidence Group (HBDSEG), has made significant steps towards achieving this aim through the following objectives:

1. Standardisation of data protocols and submission requirements across cetacean data providers and data receivers
2. Development of a data policy regarding terms of data submission, access and use.
3. Scoping of options for hosting the JCDP resource.
4. Support major NGO data providers to prepare recent data for submission to the resource
5. Create a project web page to promote the project objectives and the opportunities of a collated dataset.

A workshop was held to bring together relevant UK stakeholders to progress the first three objectives, resulting in initial drafts of a JCDP Data Standards Protocol and Data Policy. Scoping of the host platform for the JCDP was also discussed at the workshop to establish the most appropriate option, including stakeholder confidence in how data will be stored and accessed; and capacity and functionality. The preferred option was identified, and discussions were then held to establish the potential for wider collaboration on developing and implementing an international JCDP resource in future phases of this project.

Objective four provided support to key UK based charitable data providers in processing recent data to ensure their own databases are up to date, thereby facilitating future integration of those data into the JCDP.

Finally, a webspace was set up to host information on the project including the aims and objectives, and acknowledgement of those involved. The intention is for this webspace to be developed as the project develops, to become a communication tool in support of JCDP operations, such as publicising the availability of the data; support for data providers, and publication of relevant associated outputs.

The work to date has made significant progress towards achieving the aim of a readily accessible database. This first phase has focussed on working with UK stakeholders but the ambition is to build a database that will be populated with data throughout the northeast Atlantic from a broad range of organisations; this is essential to ensure that the scale of the data available better marries with that of the wide-ranging cetacean species.

Table of contents

1. Introduction 5
1.1 Background 5
1.2 Objectives 6
1.3 Project governance 6
2. Project summary 7
2.1 Approach to delivery 7
2.2 Objective 1: Coordination of data protocols and standards 7
2.3 Objective 2: Data Policy guidelines: Data Provision, Access and Use Policy 8
2.4 Objective 3: Scoping options to host the JCDP 9
2.5 Objective 4: NGO data update 12
2.6 Objective 5: Project website 13
3. Conclusions 13

Annex I Workshop Paper 15
Annex II Workshop report 31
Annex III Draft Data Standards Protocol 37
Annex IV Data validation evaluation 51
Annex V Draft Data Policy 55

1. Introduction

UK governments are required to report on the status of cetaceans (whales, dolphins and porpoise) under several national and international legislative instruments, including the UK's Marine Strategy[1] and OSPAR[2]. The species of interest are wide-ranging with some, such as the harbour porpoise (*Phocoena phocoena*), occurring throughout continental shelf waters. Estimates of population size are derived from shipboard and aerial survey data and the UK currently supports decadal surveys (Hammond et al. 2002, 2013, 2017) that cover its national waters. There are several smaller scale surveys that occur in the interim, collected by a wide variety of initiatives, which provide finer scale information on the seasonal distribution and relative abundance of cetaceans.

The decadal cetacean surveys are designed to provide robust estimates of abundance and long-term trends. However, as they are relatively infrequent, they are unable to inform on shorter-term changes in populations or on finer scale distribution. The UK assessment and reporting requirements usually occur in 6-yearly cycles and therefore there is a need to process all available data to ensure robust assessments can be made.

Currently, there is no mechanism to readily access the wide range of existing cetacean monitoring datasets and the process to do so can be labour intensive and inefficient. Also, as the datasets are collected by a wide variety of organisations and standards of data collection across them differ, analysis of a collated dataset is challenging. With this in mind, the Joint Cetacean Data Programme (JCDP) was established with the aim of addressing these issues. Ultimately, the goal of the JCDP is to create an accessible web-linked resource of available data in a database; options for how this should be delivered were also scoped out in this part of the project. Phase one of this work is outlined in this report.

1.1 Background

In 2003, JNCC published the Atlas of Cetacean Distribution in North-west European Waters[3], the result of a collation of three main sources of data (the European Seabirds at Sea (ESAS) database, Sea Watch Foundation, and Small Cetacean Abundance in the North Sea (SCANS) data). The Joint Cetacean Protocol (JCP) developed from this collaboration, with the following aims:

- Provide cetacean summary information via a web-based portal, including species-specific estimates of cetacean density, distribution and population trends
- Create a standard structure for sharing cetacean sightings data
- Allow portal users to request access to source data, while leaving their provision at the discretion of each contributing organisation
- Assist with reporting on cetacean conservation status to various Directives including the EU Habitats and Species Directive and Marine Strategy Framework Directive

These aims were met with varying degrees of success. However, over four phases of analysis (e.g. see Paxton et al. 2016), the aims of producing a data standard (for submission for analyses), production of summary information and provision of information for reporting were met. However, the aim of providing a portal from which data could be requested for third party use was not specifically delivered.

The Marine Ecosystems Research Program (MERP) undertook a more recent collation of seabird and cetacean datasets from north-west European seas, but with a wider temporal

[1] https://moat.cefas.co.uk/introduction-to-uk-marine-strategy/

[2] https://oap.ospar.org/en/

[3] http://data.jncc.gov.uk/data/a5a51895-50a1-4cd8-8f9d-8e2512345adf/atlas-cetacean-distribution-web.pdf

scope, and incorporating many datasets that were not available for JCP analyses (e.g. cetacean data from other European groups; from industry bird surveys that were not in the ESAS database; newer datasets that had not been fully processed when JCP analyses started, and over a broader geographical area incorporating data from continental European research groups.).The major outputs from the MERP included monthly cetacean densities for the period 1985 and 2017 (Waggitt et al. 2019), and since then, further datasets have been received from several research groups. However, the data collated under the MERP are not freely available, although it can be used to identify data owners so that data use permissions can be sought.

This work demonstrated the value of collating cetacean datasets, with outputs contributing to, for example, the identification and designation of Marine Protected Areas for harbour porpoise, and in Scotland for minke whale and Risso's dolphin. As a result, this project develops the concept to become a persistent resource on which to base high-quality analyses in support of UK and international monitoring, reporting and assessment needs.

1.2 Objectives

The overarching aim of the JCDP is to coordinate how cetacean data are collected, stored, accessed and analysed, for the benefit of cetacean conservation. The objectives of this phase of the project were to:

1. Standardise data protocols and submission requirements across cetacean data providers and data receivers

2. Explore options for permissions and long-term data access across data providers

3. Scope the platform options for continued collation and hosting/archiving of the cetacean dataset

4. Support major NGO data providers to prepare recent data (particularly from 2016 onwards) for submission to the resource

5. Create a project web page for promoting the project objectives and the opportunities of a collated dataset.

1.3 Project governance

A steering group was convened for the JCDP comprising key organisations from the relevant stakeholder groups (predominantly UK at this stage) including data providers, users and archivers. This included the following organisations:

Organisations	
Joint Nature Conservation Committee (JNCC) (SG chair)	Natural England
Bangor University	Marine Scotland Science
Sea Watch Foundation	Welsh Government
Centre for Research into Ecological and Environmental Modelling	ORCA
Hebridean Whale and Dolphin Trust (HWDT)	RenewableUK
Natural Resources Wales (NRW)	HiDef
Department for Agriculture, Environment and Rural Affairs (DAERA)	Plymouth University
Scottish Natural Heritage	Exeter University

Scientific Unit, National Parks & Wildlife Service	Oil and Gas UK
Marine Biological Association (MBA)	Royal Haskoning
UK Hydrographic Office (UKHO)	Hartley Anderson
Sea Mammal Research Unit (SMRU)	MARINELife
SMRU consulting	Wildlife Trusts
RPS Group	University College Cork
Galway-Mayo Institute of Technology	Cornwall Wildlife Trust
Whale and Dolphin Conservation (WDC)	NIRAS
Irish Whale and Dolphin Group (IWDG)	APEM
Scottish Power Renewables	

2. Project summary

Below is a summary of the delivery of phase one of the JCDP:

2.1 Approach to delivery

Engagement with JCDP stakeholders has been an essential part of developing the JCDP. To support this the steering group and where relevant, additional stakeholders have been consulted in progressing the objectives.

Much of the stakeholder input has been done remotely, through review of drafted information and proposals. A two-day workshop was also hosted by JNCC and involved focussed sessions on objectives 1-3 to address some of the more significant potential barriers to achieving the aim of the JCDP.

The workshop brought together a cross-section of the identified stakeholders. A workshop paper (Annex I) was produced in advance to provide contextual information and an overview of relevant information relating to the three objectives for discussion: defining a data standard; scoping a host platform and; developing a data policy. Breakout groups were integrated into the workshop to facilitate focussed discussions on key elements. Membership of these groups included participants representing the different stakeholder types wherever possible (e.g. NGO, industry, academic, public body) to ensure all views were represented in each discussion.

The workshop resulted in positive outcomes and proved a productive method of collating the relevant views of stakeholders to enable progression of the work. A workshop report was produced, capturing the discussions, outputs and actions from the sessions (Workshop).

Ad hoc meetings were also arranged with stakeholder groups as required, to discuss specific aspects of the work that needed further consideration. These discussions have fed into the project outputs as and where relevant.

2.2 Objective 1: Coordination of data protocols and standards

Establishing a data standard will make it easier for data collectors and owners to contribute to the JCDP. It will communicate the need for core data fields as a prerequisite for data submission, as this will ensure a level of data quality so that key analyses (e.g. on distribution and abundance of cetaceans) can be undertaken.

Achieving standardisation of cetacean data collection is underpinned by good relationships with the data collectors. Consequently, this objective was primarily delivered as part of the stakeholder workshop outlined in Section 2.1. In advance of the workshop, a “Data Fields and Vocabulary” matrix was developed to assist in understanding data collection by the various

organisations. This matrix was filled out during the workshop and followed up with a discussion around similarities and barriers to a standardised approach, as well as the willingness of data collectors in adapting their existing protocols to align with a JCDP standard. The matrices and discussion outcomes were then used to identify the JCDP core data fields and form the basis of a data standard for cetacean data collection. The draft JCDP Data Standards Protocol was developed and can be found in Annex III.

Further work is needed with stakeholders to try and resolve issues around adoption of the JCDP data standard. A common remaining concern was with regards to the resource needed to adapt data collection and/or transform data formats so that submission to the JCDP is possible. Eventually, some format conversions may be possible through inbuilt tools of the host platform or a JCDP coordinator (if one were to be appointed) could facilitate data preparation. This would not only involve conversion of data into the right format (e.g. changing measurement units) but also some degree of data validation and cleaning.

It was recognised that tools to help organisations to validate their data may already exist. The project undertook a comparison of methods that were applied for three at sea data-collation exercises: Marine Ecosystems Research Programme (MERP), Joint Cetacean Protocol (JCP) and European Seabirds at Sea (ESAS). A summary of the evaluation is given in Table 1 and a full report in Annex IV.

2.3 Objective 2: Data Policy guidelines: Data Provision, Access and Use Policy

The JCDP strives for open access data so that it can be available to support a wide range of cetacean-related research, policy and management needs. However, it was acknowledged that organisations may have reservations to supplying data under open access conditions. Therefore, this objective was progressed through the stakeholder workshop. The workshop discussions focussed on the nature of access to data that would eventually be held in the JCDP database. In the paper produced to support facilitation of the workshop (Annex I), a list of options for accessing data held in the JCDP was provided for discussion at the workshop. The options outlined the different ways that access to data could be managed, including e.g. open access for all; on a request basis; tiered access etc. These discussions are captured in the workshop report (Workshop).

The workshop outputs were used to develop a draft Data Policy (Annex V). The policy aims to capture stakeholder views with regards to data access but also subsequent use of the data. NGOs had particular concerns around this aspect, and further meetings were held to understand their level of comfort with regards to how their data would be treated. Industry data are often commercially sensitive during the consenting process but are routinely submitted to The Crown Estate for storage after that time. A follow-up discussion was had with a number of offshore wind developers to understand their concerns around submission and access to the JCDP. Dialogue with stakeholders will need to be maintained; the Data Policy is a draft and will undoubtedly need to be revised as the JCDP progresses.

Table 1 Summary of the evaluation of validation routines undertaken for three data collation exercises by the Marine Ecosystems Research Programme (MERP), Joint Cetacean Protocol (JCP) and European Seabirds at Sea (ESAS). Those collecting and providing data are referred to as suppliers whereas those collating data are referred to as collators. Manual validation describes checks achieved by human inspection of graphs, maps and spreadsheets. Automated validation describes checks achieved by computer algorithms.

Process	MERP	JCP	ESAS
Manual Validation?	Yes. Performed by collators.	Yes. Performed by suppliers.	Yes. Performed by collators.
Automated Validation?	Yes. Performed by collators.	Yes. Performed by collators.	Yes. Performed by collators.
Missing Information Identified and Sought?	Yes. Manual Validation.	Not stated.	Yes. Identified in Automated Validation and checked in Manual Validation.
Search for Duplicated Rows?	Yes. Automated Validation.	Yes. Manual Validation.	Yes. Identified in Automated Validation and checked in Manual Validation.
Invalid Platform Speeds Removed?	Yes. Automated validation. Based on mean recorded speed per platform-type and supplier.	Yes. Automated Validation. Based on maximum theorectical speed per platform category	Yes. Manual Validation. Based on data visualisation.
Effort on Land Removed?	Yes. Automated Validation. Interpolation between waypoints, with routes on land removed.	Yes. Manual Validation. Based on data visualiation.	Yes. Manual Validation. Based on data visualisation.
Sightings without associated effort removed?	Yes. Automated Validation. Sightings that occur outside time-limits of validated sections are removed.	Yes. Automated Validation. Sightings that occur outside time-limits of validated sections are removed.	Yes. Identified in Automated Validation and checked in Manual Validation.
Sightings on Land Removed?	Yes. Automated validation. Those situated on land are removed.	Yes. Manual Validation. Based on data visualisation.	Yes. Manual Validation. Based on data visualisation.
Sightings too far from transect line (i.e. incorrect coordinates) removed?	No.	No.	No.
Manual "sense-checks" using literature and expert knowledge?	Yes.	Not stated.	Not stated.

2.4 Objective 3: Scoping options to host the JCDP

The JCDP requires a host platform for the database and support a user-friendly front end for access to the datasets. There is a strong desire to tie the JCDP in with existing examples where feasible, to ensure the JCDP integrates with what is already available regarding marine data storage and sharing.

In order to identify the best way forward, a summary of existing platforms and other relevant resources was compiled within the workshop paper (Annex I) to inform discussion at the workshop. Presentations at the workshop provided an overview of some of the existing “data collation” platforms, for example on The Archive for Marine Species and Habitats Data (DASSH) and the European Seabirds at Sea (ESAS) database. These examples were provided for stakeholders to support discussions on the most appropriate way forward for the

JCDP and to consider whether an existing platform may meet the project's needs or whether a bespoke system for the programme needs to be developed. The focus at this stage was on developing a UK resource, bearing in mind the desire to enable the JCDP to encompass data collected over the wider North-east Atlantic region in future, given the highly mobile nature of marine mammals.

An overview of the discussions can be found in the workshop report in Annex II. The overall view was that collaborating with the ICES data centre was the most suitable way forward to pursue, given their international remit and independence from any single country, which would make it a more attractive prospect for buy in from non-UK data collectors. ICES have extensive resource and experience in developing these types of platforms and they have a prominent position in the scientific community which would enable immediate recognition and wider publication of the JCDP resource.

Subsequently, a meeting was set up with representatives from the ICES data centre, to establish what was possible with regards to ICES building and hosting the proposed JCDP resource. It was confirmed that development of the JCDP within the ICES data centre was a possibility, and we identified the various routes and costs to achieve it.

ICES offer a very cost-effective system for collating, storing, maintaining, and providing some standard analyses of cetacean data. The access structure that ICES could deliver, however, would need to be relatively simple and align with ICES data policies.

JNCC were conscious of stakeholder discussions with regards to levels of access permissions and expectations in terms of financial contributions in support of contributing data. There was also benefit in understanding better the volume of potential data contributions to the JCDP from the different "sectors" of data collectors, how this has changed over time and how data gaps may be tackled through maximising the availability and quality of data e.g. year-round data. Therefore, some analyses were carried out on the MERP dataset (1980-2018) as it is the largest collation of cetacean data to date for its UK data owners to demonstrate that all sectors can potentially make valuable contributions to the JCDP, especially when examined relatively across time periods and seasonally.

An overview of data contributions is illustrated below in the tables. Table two shows that NGOs have become the dominant single source of cetacean data collection in the UK over recent years although collectively UK government funded initiatives have contributed the bulk of data and coverage over the last three decades. Seasonally, the commercial sector is a valuable data provider during winter months (Table 3). Over the decades, NGOs have had a growing contribution of data and have recently become the dominant source of cetacean data collection in the UK (Table 4). It should be noted that SCANS III and some recent NGO data were not included in MERP as it was not prepared in time for the analyses undertaken in the project. However, the tables still offer a representative overview of the data contributions from each sector.

Table 2 Data submitted by UK data owners to the MERP data collation exercise 1980-2018 by sector. European Seabirds At Sea (ESAS) contains a collation of data from Academic, Commercial, Overseas and UK Government which could not be ascertained from the database. For transparency, ESAS has been treated separately to others where the original source is known.

Source	Km	% Contribution
Academic	17267	1
Commercial	189902	16
ESAS	427929	37
Overseas Government	26270	2
Overseas NGO	46672	4
UK Government	109531	9
UK NGO	342535	30

Table 3 Data submitted by UK data owners to the MERP data collation exercise 1980-2018 by sector and seasons. European Seabirds At Sea (ESAS) contains a collation of data from Academic, Commercial, Overseas and UK Government which could not be ascertained from the database. For transparency, ESAS has been treated separately to others where the original source is known.

Source	% Contribution			
	Dec-Feb	Mar-May	Jun-Aug	Sep-Nov
Academic	1	1	2	2
Commercial	34	18	10	17
ESAS	49	40	31	39
Overseas Government	3	1	3	2
Overseas NGO	5	5	3	5
UK Government	2	5	14	9
UK NGO	7	30	38	26

Table 4 Data submitted by UK data owners to the MERP data collation exercise 1980-2018 by sector and decade. European Seabirds At Sea (ESAS) contains a collation of data from Academic, Commercial, Overseas and UK Government which could not be ascertained from the database. For transparency, ESAS has been treated separately to others where the original source is known.

Source	% Contribution			
	1980s	1990s	2000s	2010s
Academic	0	0	1	4
Commercial	0	2	37	13
ESAS	99	74	6	2
Overseas Government	0	3	1	4
Overseas NGO	0	0	7	7
UK Government	0	3	5	28
UK NGO	1	18	43	41

Discussions at the workshop also touched on options for financial return to data contributors (i.e. data downloads would be chargeable for some uses/users) and as a means of generating funds to support the JCDP and its data providers long-term. Discussion around how this could be implemented into a host platform quickly became complex and it was agreed not to pursue this under the current phase of work. Such a condition on data download would also not be agreeable to ICES. As such, this needs to be considered when deciding on options for hosting the database. However, the NGO sector expressed the issues they face to sustain long-term data collection, which is reliant on volunteer surveyors, partnerships with ship-owners (e.g. ferry companies) and grant money to support staff roles and training. Rough estimates of annual costs from three UK NGOs to support their data collection efforts are in the region of £200 - 300k each.

2.5 Objective 4: NGO data update

Cetacean data are collected by a variety of different providers, with a variety of different means. The NGO sector relies on fundraising and/or volunteers to collect data. The ability to fund collection and processing is therefore an ongoing issue for NGO providers, yet their data resource would be an invaluable contribution to the JCDP. As a result, key NGO data providers were identified based on the quantity of data that they have contributed historically to the JCP and MERP projects. The JCDP recognised that some funding to support data processing (at least 2016-2018) would ensure that should the relevant NGOs decide to contribute data to the JCDP database, then their data holdings are relatively up to date. Four NGOs: ORCA, Hebridean Whale and Dolphin Trust, Sea Watch Foundation and MARINELife, were provided with some funding within phase one of the project for this purpose. These organisations processed their most recent data (e.g. transcribing from paper sheets to digital format) and supplied the project with comprehensive metadata, including their survey protocols.

Under this objective, HWDT was able to process and validate all survey and sightings data from their survey yacht *Silurian* for the period April 2016 to December 2019. This consisted of 32,036 kilometres of visual and/or acoustic survey effort and an associated 4,591 sightings of marine megafauna (cetaceans, pinnipeds, sharks).

MARINELife were able to enter and validate all data collected, except for some small boat surveys off north east and south west England. All ferry, freight and research cruise data were updated and selected small boat data amounting to a total of 327 surveys and adding a combined total of 116,413 km of survey effort. These data included 2,854 cetacean sightings.

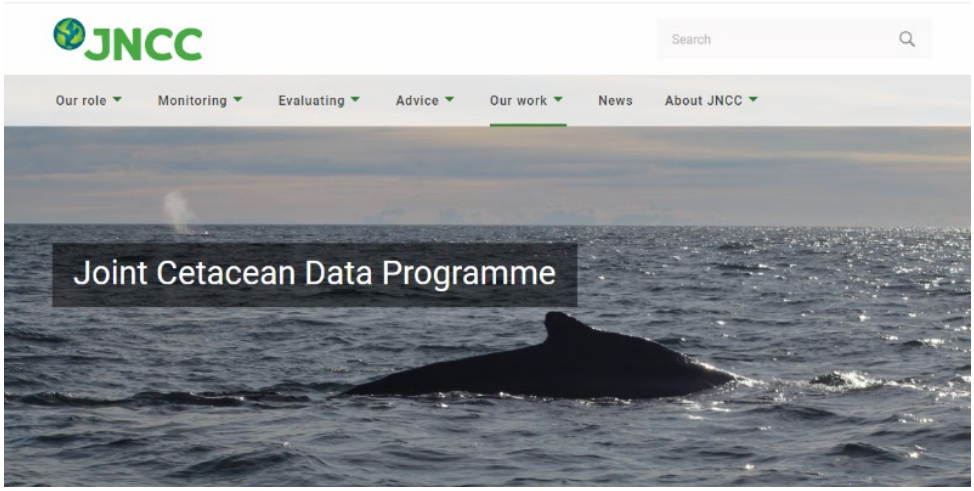
ORCA has cleaned and validated all distance sampling survey effort and sightings data for the period 2016 to 2019. This consists of 302 surveys across 13 different ferry routes and equates to a total 78,695.5 km of effort. An associated 3,092 sightings of marine megafauna (cetaceans, sharks, fish and seals) were recorded, consisting of 12,256 individual animals.

Sea Watch Foundation has processed, cleaned and validated 82,095 km of survey data for the period 2015 to 2019. The majority of these were double-platform line transect data. An associated 4,779 sightings of marine megafauna (largely cetaceans but also some pinnipeds and sharks) were recorded.

2.6 Objective 5: Project website

A JCDP webpage was set up at the start of the project as a place to communicate the objectives of the work and acknowledge those involved. The webpage can be found here:

https://jncc.gov.uk/our-work/joint-cetacean-data-programme



The page is a simple overview of the first phase of the project, pending the outcomes of Objective 3 – scoping options to host the JCDP, which will inform the development of the JCDP online presence. As the project lead, JNCC currently hosts the webpage and it has the following contents:

1. Background
2. Year 1 project aims
3. Project funders
4. Project partners
5. Project Steering Group

3. Conclusions

Phase one (2019 – 2020) of the JCDP has delivered on all five of the objectives:

1. *Standardise data protocols and submission requirements across cetacean data providers and data receivers*

First draft of the Data Standards Protocol complete (Annex III)

2. *Explore options for permissions and long-term data access across data providers*

First draft of the Data Policy complete (Annex V)

3. *Scope the platform options for continued collation and hosting/archiving of the cetacean dataset*

ICES identified as the primary option for progressing the development of the JCDP platform. They can offer design, maintenance, access and update solutions and as an international organisation, promotes the JCDP as an international resource.

4. *Support major NGO data providers to prepare recent data (particularly from 2016 onwards) for submission to the resource*

Funds provided to four NGOs to support processing of their most recent data in readiness for collated use. Data collected from 2016 to 2018, and some more recent, have been processed, and the metadata provided to JNCC.

5. *Create a project web page for promoting the project objectives and the opportunities of a collated dataset.*

JCDP webspace currently hosted on the JNCC website, hosting information on the ultimate project aim; phase one of the funding; and promotion of project collaborators.

The work to date has made significant progress towards achieving the aim of creating an accessible and comprehensive database of UK cetacean data. The Data Policy and Data Standard remain as “draft” given that these will likely need to be updated as the project progresses, and the host platform is designed and built. For example, if ICES were to build the platform it may be efficient for the cetacean datasets to align, where appropriate, with their existing data structures and vocabulary. Further funding will need to be sought to realise the database and user-friendly platform for data access and submission. JNCC will progress the effort with support of the PSG and wider stakeholders. Maintaining the momentum achieved under this first phase of work is essential to the eventual success of the JCDP. If the project can be realised within the next 12-24 months then these data and associated products can be utilised for a range of future assessment and reporting commitments, such as the next iteration of OSPAR and Marine Strategy assessments.

Annex I Workshop Paper

This paper was prepared and circulated to the JCDP steering group and workshop attendees in advance of the workshop, to support productive discussion:

Joint Cetacean Data Programme (JCDP)

Stakeholder Workshop Discussion Paper

Working towards a data standard, data access agreement and host platform solution

1 Aim of the Joint Cetacean Data Programme

Mobile marine species such as cetaceans, inhabit an environment which is challenging to survey and observe, which limits understanding of their complex ecology and biology. We rely on evidence to inform decisions on sustainable sea use; welfare; population status and emerging and changing pressures such as climate change. Decisions still need to be made even in the absence of a good evidence base, but a lack of understanding has an impact on the effectiveness of decisions on how we manage our seas for the benefit of all ecosystem functions.

The Joint Cetacean Data Programme (JCDP) aims to provide a growing, collated dataset of cetacean survey data to serve as a resource for stakeholders to perform a suite of analyses to inform conservation, research and policy needs.

2 Background

In 2003, JNCC published the Atlas of Cetacean distribution in north-west European waters, the result of a collation of three main sources of data (the European Seabirds at Sea (ESAS) database, Sea Watch Foundation, and Small Cetacean Abundance in the North Sea (SCANS) data). The Joint Cetacean Protocol developed from this collaboration, with the following aims;

- provide cetacean summary information, via a web-based portal, including species specific estimates of cetacean density, distribution and population trends;
- create a standard structure for sharing cetacean sightings data;
- allow portal users to request access to source data, while leaving their provision at the discretion of each contributing organisation;
- assist with reporting on cetacean conservation status to various Directives including the EU Habitats and Species Directive and Marine Strategy Framework Directive.

These aims were met with varying degrees of success. However, over four phases of analysis[4] (e.g. see Paxton et al. 2016), the aims of producing a data standard (for submission for analyses), production of summary information and provision of information for reporting were met. However, the aim of providing a portal from which data could be requested for third party use was not specifically delivered.

The Marine Ecosystems Research Program (MERP) undertook a more recent collation of seabird and cetacean datasets from north-west European seas, but with a wider temporal scope, and incorporating many datasets that were not available for JCP analyses (e.g. cetacean data from other European groups, from industry bird surveys that were not in the ESAS database, and newer datasets that had not been fully processed when JCP analyses started).The major outputs from the MERP included monthly cetacean densities for the period 1985 and 2017 (Waggitt et al. 2019), and since then further datasets have been received from several research groups.

This preceding work has demonstrated the value of collating these data, with outputs contributing to, for example, the identification and designation of Marine Protected Areas for harbour porpoise, and in Scotland for minke whale and Risso's dolphin. As a result, this project will develop the concept to become a persistent resource from which to continue high-quality analyses and assessments.

[4] https://webarchive.nationalarchives.gov.uk/20180804195742/http://jncc.defra.gov.uk/page-5657

3 Defining a data standard

3.1 Potential data applications (non-exhaustive)

The effective assessment of highly mobile species often requires extensive datasets with large geographical and temporal scope. However, within the UK, the entirety of national waters has only been surveyed at decadal intervals as part of the Small Cetacean Abundance in the North Sea and European Atlantic (SCANS) projects. By collating smaller-scale surveys throughout the UK, and over time, a long-term dataset with coverage through UK waters can be better achieved and stored within a single database. Some of the applications of such a dataset would include:

- UK-level and international reporting and assessments of cetacean conservation status
- Identification of
    - areas of importance
    - persistent use for e.g. spatial measures such as Marine Protected Area identification
    - temporally discrete 'hot spots' of distribution for reactive management (e.g. bycatch risk)
    - changing and emerging threats and pressures
- Climate change impact surveillance
- Sensitivity of cetaceans to marine pressures
- Spatial and temporal analyses to support impact assessments and spatial planning
- Trend analysis at relevant spatial and temporal scales
- Evidence base to inform education and awareness

3.2 The need for a data standard

For datasets to be collated into a single database, there needs to be a commonality at least between some fields within the data. Having a defined standard for data collection will ease that process in the future. Standardising data offers advantages in four main areas: development of expertise and data quality, suitability of data for analyses, ease of data ingestion, and compatibility. These components all contribute towards maximising the use of independent datasets.

3.2.1 **Development of expertise and data quality**

Where field data collection is standardised, surveyor expertise increases due to familiarity with an agreed, robust protocol. As a result, data quality improves through coordinated use of an appropriate survey methodology, as well as improved ability of observers to carry out accurate and effective surveys.

3.2.2 **Suitability of data for analyses**

Standardisation of data at collection stage ensures that the requisite parameters are collected using the correct methods to do so. It also ensures that the data are recorded using the correct naming convention, taxonomy, and other associated coding. Recording supporting information e.g. spatial and environmental data will also be controlled, ensuring these data are suitable for use either independently, or collaboratively alongside other datasets.

3.2.3 **Data ingestion process**

Standardisation of datasets allows for automated upload and validation of data, saving time for both data owners and those hosting the data. The validation ensures that only those data that meet the standard will be stored, helping to maintain the quality of the data within the database.

3.2.4 Compatibility of datasets for combined use

Standardised data ease the burden of standardisation on those using the data. Data may need to be adapted to be used in analytical packages such as 'R' or 'Distance' but applying these adaptations from collection rather than to individual datasets further along the line, speeds up the process, eases data processing congestions and reduces opportunity for error.

3.3 Data collection protocols

Data standards can be met through the development of a robust data collection protocol that sets out the data collection methodology to be used. There are many existing data collection programmes for cetacean distribution and abundance data which all have established protocols to follow, but with differing standards and stipulations as to how to collect, store and/or submit data to relevant organisations and repositories. In developing a standard, there is a need to evaluate existing protocols and find a solution to creating a protocol which is fit for purpose, but with as little burden on long-standing data collection initiatives as possible to be able to contribute. In the UK, a range of methodologies have been employed to collect monitoring data that are amenable to investigating abundance, distribution and temporal/spatial changes in these parameters.

3.3.1 Double-platform line transect shipboard surveys

The protocol used for the Small Cetaceans in European Atlantic waters and the North Sea (SCANS) surveys exemplify this methodology (Hammond et al., 2002; Hammond et al. 2013). SCANS shipboard surveys are double-platform (primary and tracker) observer surveys. Two teams of experienced observers search a 180° arc ahead of the vessel with naked eye (primary platform) or binoculars (tracker platform). Distance sampling data are recorded, and duplicate detections are identified in the field by the "duplicate identifier"; duplicate sightings are used to correct for animals missed on the transect line. The protocol is designed to result in robust data suitable for estimating absolute abundance. Typically, it is undertaken aboard large vessels although it is also carried out on some smaller vessels which are suited to the methodology.

3.3.2 Single-platform line transect surveys

This methodology is commonly used by surveyors on small boats that cannot accommodate a double-platform set up, and by volunteer survey networks operating from vessels of opportunity.

For example, the European Cetacean Monitoring Coalition (ECMC) is a collaborative of NGOs with similar survey methodologies, which have come together to collate data for use as a combined dataset in favour of high-quality research and evidence. The majority of these surveys are conducted from platforms of opportunity such as ferries. Surveys are single-platform effort-related and 1-2 observers search ahead of the vessel in a 90/180-degree arc. Distance sampling data (radial distance and bearing) to each observation encountered are generally recorded and scanning is done with the naked eye and binoculars by each observer. Observers are trained in methods and cetacean identification by the respective organisation, predominantly in a classroom, and then placed with experienced surveyors for development. The data structure, coding and storage is in line with ECMC standards[5] as defined by the project. The data collected are primarily suitable for analyses of relative density rather than absolute because they cannot be corrected for animals missed on the transect line.

Cetacean data generated through offshore industry surveys will encompass several different survey methods and take a variety of formats. Some likely are single-platform line transect surveys for cetaceans, whilst others may be dual-purpose that employ a protocol for seabird and cetacean data collection simultaneously (see 3.3.3*)*. Data from offshore windfarm surveys

[5] http://www.ecmcweb.org/

are currently submitted to The Crown Estate (TCE) but not, necessarily, accompanied with the protocol used to collect the data. Data are submitted via an online portal, which ensures that they meet TCE's structural and metadata requirements, which are MEDIN[6] compliant.

3.3.3 European seabirds at Sea (modified strip transect)

The early development of the European Seabirds at Sea (ESAS) methodology (Tasker et al. 1984) was based on a strip transect approach. Strip transects involve observers surveying and making detections within a fixed width "strip" and the assumption is that all objects (e.g. birds) are detected within that strip. However, where detection of the objects decreases with distance from the survey platform, a line transect method is preferable. The ESAS methodology thus developed to account for imperfect detection. For ESAS surveys, the strip is typically 300m wide (one side of the "line") and was latterly subdivided with distance bands within it. Recording of detections by distance bands enables detection functions to be fitted. However, focus is within the 300-metre strip but detections from beyond this are recorded but without additional distance information. Data collected on ESAS surveys are primarily on seabirds but include observations of cetaceans (and other mammals, fish etc). Data are collected from a 90-degree arc (i.e. one side of the survey platform), with detections restricted to those made with the naked eye.

3.3.4 Circle-back (or "race-track") aerial surveys

This distance sampling method was used during the SCANS aerial surveys (Hammond et al 2013 & 2017) and is used so that absolute abundance estimates can be generated having been corrected for animals missed on the transect line. In this approach, on detecting a group of animals, the aircraft circles back to resurvey a defined segment of transect (Hiby, 1999); the full protocol is available in Gilles et al (2009).

3.3.5 Digital aerial surveys

The use of digital video and still cameras instead of observers on aerial surveys has steadily increased over the last decade. The resolution of the camera systems varies, and the sampling is determined to some extent by the objective of the survey. Companies that operate these surveys have survey protocols in place.

3.4 Proposed essential data fields and vocabulary

There are core data fields which are essential to enable effective use of the data for many applications (e.g. species, locations), and others which are not necessarily essential to answer key questions and complete primary analyses (e.g. observer name). The JCDP aims to coordinate the recording of, particularly core fields, for cetacean data collection at sea. The aim is not to dictate significant changes in how data are collected, particularly for established projects, but to provide a JCDP 'standard' to guide how data are recorded for the reasons outlined in Section 3.2. The JCDP builds on the protocol established by its predecessor, the Joint Cetacean Protocol (JCP) (see Appendix I); the Protocol sets out the fields required for data entry to that project but could be improved upon with regards to provision of a standard data-recording vocabulary. Other monitoring or data archive organisations have also developed vocabularies associated with cetacean monitoring datasets which include the ESAS coding manual; Sea Watch Foundation database manual, ECMC data dictionary and the Darwin Core Schema.

[6] https://www.oceannet.org/

3.4.1 **Essential metadata and format**

There are certain data fields that are essential in order to identify the survey.

Field	**Description**
Date	Date of the first day of the survey? Last day? Or range of dates?
Platform type	What was the survey platform
Vantage point	The location of the observer on the platform
Platform height	The (eye/platform) height of the observer above sea level (metres) (ship-based) Flight height (aerial)
Origin (i.e. who data belong to)	The organisation that owns the data
Survey type	The survey methodology (see Section 3.3)

3.4.2 **Effort**

Effort is essential for *certain types* of analysis and is therefore recommended as part of a survey protocol where possible to enable maximum potential of the data. However, data without effort records may still be used in basic analysis of e.g. distribution.

Field	**Description**
Record type	What is the record stating? Change of observer? Change in environmental conditions? Etc.
Time	hh:mm start and end time of survey effort
Speed	Km per hour
Angle of view	Search angle of observer (e.g. abeam port to starboard 10 degrees)
Primary search method	How was the search primarily completed? E.g. naked eye; binoculars
Latitude	Start/end decimal degrees
Longitude	Start/end decimal degrees

3.4.3 **Environmental**

The environmental conditions whilst surveying have an impact on probability of observing animals and is therefore essential information to record alongside observations.

Field	**Description**
Sea state	Sea state using the Beaufort Scale
Swell/wave height	Metres
Glare	Location of search area affected by glare

3.4.4 **Observation – species, number, behaviour**

Accurate recording of a sighting is essential to the dataset.

Field	Description
Species	Pre-defined species codes
Number	Number of individuals of a species in each sighting
Range	Distance of animal(s) in metres from the observer
Bearing	Angle of sighting from the observer
Observer	Full name
Confidence	Confidence in species id

4 Data access proposals

4.1 The need for centrally stored, accessible data

By collating available datasets, we are able to increase the evidence base significantly and work towards an improved understanding and in turn, develop well-informed management and conservation strategies to ensure healthy, sustainable populations which fulfil their functions as part of the wider ecosystem. Ideally, these data would be publicly available so all analyses could be performed using the same evidence base, increasing value and application of the outputs. Publicly funded data are becoming increasingly openly available throughout the world, released by governments as part of the transparency agenda. Where data collection is not publicly funded, there is a need to work towards as open a policy as possible to maximise the potential of these data in increasing the evidence base.

The decadal frequency of publicly funded large-scale cetacean surveys does not enable the analysis of cetacean population trends within the timescales required for assessments of status of cetaceans (e.g. every 6 years) or to underpin management decisions. Annual and seasonal surveys at a national scale would be costly so there is a need to maximise the potential within all data collected at smaller spatial and temporal scales by a variety of projects. Individually, these data from discrete projects have limited application when considering a highly mobile species. However, it has been demonstrated that, when combined, the evidence base can become a resource which allows confidence in the analyses and subsequent outputs (e.g. Paxton et al., 2016; Waggitt et al., 2019).

Therefore, if the aims of the JCDP are realised, the UK would have the benefit of a data resource to enhance our understanding of, and approach to cetacean conservation.

4.2 Benefits of central storage

In the first instance, the JCDP would act as storage for data providers, reducing the burden on individual storage systems and enabling access to your own data as and when required. If comparable data are collated in a single location, there is opportunity to develop access agreements and reduce the need for repeated permission requests for use of data. This results in a stable dataset that can be used for repeat analyses, without the need for repeat collation and permission-seeking exercises, which are time consuming, costly and often result in different sets of data depending on which data are accessible for each event.

Use of datasets in collaboration with others maximises the value of data collection efforts from individual projects which may be constrained either by spatial coverage or available effort, which ensures the resource expended on data collection contributes effectively to management of highly mobile species.

The UK would have a world-class resource, making the best use of all comparable datasets collected by a wide variety of initiatives and projects, ensuring these data are used for the shared goal of high-quality evidence-based decisions for cetaceans using our waters.

4.3 Considerations for data sharing, and proposed solutions

Although there are clear benefits to sharing data, there are considerations which must be addressed to ensure data are treated as the owner permits.

4.3.1 Data protection for personal data

The General Data Protection Regulation (GDPR) necessitates the correct processing and storage of personal data. Where personal data such as surveyor names are provided alongside records, these would need to be removed or anonymised in the database in line with the Regulation. JNCC has policies in place to ensure GDPR is applied to all aspects of work where personal data are handled.

4.3.2 Freedom of Information Act and Environmental Information Regulation

Requests for information held by public bodies are handled under the following legislation and their purpose is to make public bodies, such as JNCC, transparent and accountable.

> Freedom of Information Act 2000[7] – *gives the public a right of access to information held by all public authorities in the UK.*
>
> Environmental Information Regulations 2004[8] – *gives the public the right to access environmental information held by public authorities.*

The right to this information is subject to certain exemptions and exceptions.

4.3.3 Risk of loss of data ownership

Submitting data to be held outside of the contributor organisation holds a risk that control of those data is lost and ownership is no longer with the data provider. Submission of data to the JCDP would be enabled under an agreement which will detail how the data will be treated. Ownership of the data would remain with the provider and use of those data controlled under the data access agreement that will be developed with stakeholders as part of this project.

4.3.4 Value of data to contributing party

Publicly-funded data are the property of the public, which results in them being made publicly available either explicitly via various platforms, or on request. Data collected via other funding means hold value for the contributor either in monetary terms, and/or in support of their organisational objectives. In order to support access to data to achieve the objectives of this project, there is a need to work with data providers to find an appropriate solution. This would take the form of a data access agreement, which stipulates the terms under which the data may be used.

4.3.5 Perceived misuse of data

The term ‘misuse’ can be applied to two separate types of action. In some cases, data are accessed and used in good faith, but with inappropriate analytical methods or other processes. This type of misuse can be mitigated by the provision of broad usage guidelines or detailed descriptions of how the data might best be used, and where its limitations lie. JCDP

---

[7] https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/freedom-of-information-act/

[8] https://ico.org.uk/for-organisations/guide-to-the-environmental-information-regulations/what-are-the-eir/

data should be supplied with robust guidance on its application, and if possible, JCDP expertise should be made available to those who want to use the data, to increase the value of outputs and reduce publication of errant information.

'Misuse' can also apply to cases where those using the data do not share the same values as those who supply it. For example, some organisations may object to their data being used in support of justification for e.g. commercial developments which do not align with the values of the data provider. If data are not publicly available, then this kind of misuse can be regulated by those controlling access. To satisfy data providers in submitting to the JCDP, options for a data access protocol have been outlined and input will be sought to find a way forward in achieving the project objectives, but also ensure confidence from data providers in how their data may be used.

4.3.6 **Cost of data processing and standardisation**

In order to collate data from different sources, there is a requirement to apply some level of standardisation to enable combined use. This bears a resource burden in terms of adapting collection and storage protocols to some degree. However, this is a short-term burden which enables spatially or temporally discrete datasets to be used in combination, maximising the capacity of all projects contributing to cetacean monitoring data and enabling robust analyses at relevant scales, a shared goal of many projects collecting these data.

4.3.7 **Equality**

There is risk that data providers perceive inequality with regards to what is put in to the database, and what can be used as a result. Contributors will have varying levels of data to submit and data users may or may not have contributed data to the project, depending on the data access agreement terms.

Data contribution: discussions with data providers will need to be had to find a way forward with fair treatment, taking into consideration the contributions to the database.

Data use: stakeholders will support the development of a data access agreement, which will set out the terms of use of the data held within the database.

4.4 Options for access to a JCDP dataset

A list of options has been devised drawing on experience from existing data sharing platforms, and discussions from previous meetings held in advance of funding this next stage of this work. The capacity to combine options gives a greater chance of reaching agreement between all partners. Reasonably, the resource could combine some open access data, freely available data products, as well as tiered access of some sort. However, the resource needs to be user friendly, so a solution that is relatively simple to understand and access, would be beneficial. Furthermore, the more complex the system, the more costly it will be to set up and manage, which risks the possibility of a funding shortfall to enable the resource.

Options summary

i. Fully open access to raw datasets from all contributors

This would be the simplest option to go with in terms of ease of setting up the interface for submission and extraction; in enabling a constant dataset for repeat use and; reduced need for repetitive requests for permissions. However, as described in this paper it is appreciated that there are barriers to this. It remains the ultimate objective, but the other options offer various stipulations that are designed to overcome the barriers and enable all to reach an agreement.

ii. Fully open access to data products, raw data still restricted

Processed data would be freely available in the form of agreed outputs such as distribution layers, but raw data would remain in the control of the provider for use only on request. This would restrict the uses of the data on an open access basis, and the need for users to seek permission would remain. Permission requests would need to be considered within the interface to ease the burden on all involved.

iii. Tiered access: use-related

This option would place a tier system with regards to data use, depending on an upfront agreement from data providers. For example, the full dataset could be available to an agreed set of users on an open basis (e.g. for UK status assessments), with restrictions applied for others (e.g. research projects from parties who haven't contributed data to the project; those using data for commercial benefit).

iv. Tiered access: time-related

This option would place restrictions on access to data with regards to an agreement for open access following expiration of a timeframe. For example, after an agreed amount of time to allow data providers to use data for organisational objectives; following the outcome of a plan or project application.

v. Tiered access: contribution-related

This option takes into account benefits for data providers, over those who wish to use the data but do not contribute. It may also account for level of contribution amongst providers, in order to encourage fair treatment in relation to level of contribution.

vi. Tiered access: subscription or paid access

This generates revenue for the project in order to enable e.g. further outputs to be produced. There would be a subscription, or single payment system for use of data, as agreed by stakeholders.

vii. Access on request

This would require all uses of data to be via request to data providers, which would be facilitated through the platform.

Table 1 **Data access options**provides a suite of options for access to the JCDP dataset, with suggested pro's and con's for each. The solution chosen may be one of these options, or in the case of not opting for open access, a combination of these options. Data generated by publicly funded surveys must be made open access, so any solution must accommodate this.

Table 1 Data access options

	Option	Explanation	Pro's	Con's
i	Fully open access to raw data	Raw data are openly available for download with no restrictions (GDPR compliant)	Largest data pool available for all, maximisation of use of data to contribute to understanding of cetaceans	No control over what data are used for
			Simplest, most cost-effective solution to implement	Data lose any potential commercial value to owners
				Opportunity for commercial gain from data users
ii	Fully open access to data products	Data outputs e.g. densities modelled from data at agreed temporal and spatial scale are available for download with no restrictions. Raw data would still be subject to a request to data providers.	Protects control over raw data	Objectives of analysis need to be agreed. Project scope considerably narrowed. Models need to be agreed, designed, and updated with new data
			Products maximise use of all data without making raw data available	Public data will need to be open access, so will need a system accommodating both options
			Relatively easy to implement (e.g. through ICES)	Increased cost to apply restrictions to the system
				Requirement to produce up to date outputs for use
				Cost implications of applying restrictions, and would significantly restrict use of the dataset
iii	Tiered access, use-related	Raw data are available for certain pre-agreed uses. Otherwise restricted	Some commercial value may be retained	Time requirement on each data owner
			Some control retained	Time requirement on central data holder
				Agreement needed on which uses are given access
				Some data will need to be open access, so will need a system accommodating both options
iv	Tiered access: time-related	Raw data are made available after a certain time-period e.g.	Older data become open access	Time requirement on each data owner
			Some commercial value retained	Time requirement on central data holder

		an agreed age; following a project application outcome (industry)	Some protection retained	Criteria need to be agreed for release of data
				Some data will need to be open access, so will need a system accommodating both options - cost
v	Tiered access: contribution-related	Raw data are available to those who contribute to dataset – terms to be agreed	Some commercial value retained	Bespoke/private web resource potentially needed
			Some protection retained	Potential reluctance to trade access to large datasets for access to smaller ones
			Potentially creates buy-in	Some data will need to be open access, so will need a system accommodating both options
vi	Tiered access: subscription or paid access	Raw data are available to those who pay for data or pay a subscription	Generates funds which can support the project e.g. producing outputs.	Data less likely to be used, or favours commercial use where funding is more accessible
			Potentially creates more buy-in	Some data will need to be open access, so will need a system accommodating both options
			Paying for data allows owners to retain control	Issues with how to administer the funds
			Realises commercial value of data	
vii	Unstructured access	Raw data are made available at owner's discretion case by case	Full control retained	Some data will need to be open access, so will need a system accommodating both options
				Lack of clarity on what data will/won't be supplied for
				Unlikely that analyses will have access to the same resource, depending on what permission is given to which project
				High level of admin remains, in requesting permission from multiple data providers for each use
				Not conducive to long term agreement from partners

5 JCDP platform scoping

In order for the JCDP to become a resource, the programme will need to be housed and maintained in the long-term which takes funds and human resource. The where and how needs to be established. There are various projects which exist for the purpose of making data accessible and useable, which provide helpful examples on which to base JCDP discussions.

5.1 Existing data repositories, portals or platforms

Projects exist which aim to bring together datasets and make them available and accessible for wider use. A work package of this project is to review these platforms and scope out the most appropriate way forward for hosting the JCDP 'database' to enable a stable resource, taking existing examples and standards into account. There is a need to ensure the standards and data collection guidance developed for this project, are coordinated with existing initiatives such as those listed in Table 2 **Existing data repositories, portals or platforms**, to ensure the resource is compatible with these systems.

Table 2 Existing data repositories, portals or platforms

Platform/repository	Summary
OBIS Ocean Biogeographic Information System	OBIS is a global open-access data and information clearing-house on marine biodiversity for science, conservation and sustainable development. The vision is to be '*the most comprehensive gateway to the world's ocean biodiversity and biogeographic data and information required to address pressing coastal and world ocean concerns'*. The Marine Biological Association (MBA) has managed the UK 'node' for OBIS since joining in 2018. https://obis.org/
ICES International Council for the Exploration of the Sea	ICES is an intergovernmental marine science organisation, meeting societal needs for impartial evidence on the state and sustainable use of our seas and oceans. The organisational goal is to advance and share scientific understanding of marine ecosystems and the services they provide and to use this knowledge to generate state-of-the-art advice for meeting conservation, management, and sustainability goals. Data are organised into themes, such as the biodiversity dataset which hosts seabird and seal abundance and distribution records and is linked to OSPAR, and ICES groups (JWGBIRD, WGMME). http://www.ices.dk/marine-data/data-portals/Pages/default.aspx
EMODnet European Marine Observation and Data Network	EMODnet consists of ~160 organisations that work on assembling, harmonising and making marine data, products and metadata more available to public and private users. This data ingestion portal facilitates additional data managers to ingest their marine datasets for further processing, publishing as open data and contributing to applications for society. https://www.emodnet-ingestion.eu/
MEDIN and DAC	MEDIN is a portal which is open to all with an interest in marine data and information. The data are held within dedicated DACs and linked through the MEDIN portal which stores the metadata. It is sponsored by a consortium

Marine Environmental and Data Information Network and Data Archive Centres	of sponsors and partner organisations representing government departments and agencies, research organisations and private companies and have committed to practise good data management to help future-proofing and secure the UK's marine data. MEDIN reports to the Marine Science Coordination Committee, which is a government initiated working group which aims to identify opportunities for the alignment and development of UK marine science to inform policy decisions and forward implementation of the UK Marine Science Strategy https://www.gov.uk/government/groups/marine-science-co-ordination-committee https://www.medin.org.uk/
ECMC European Cetacean Monitoring Coalition	The ECMC is Europe's largest partnership of ferry-based surveying organisations, currently consisting of several NGOs and research groups, which collect and contribute marine mammal survey data to a centralised location. The project has developed a data standard and ingestion process. www.ecmcweb.org
TCE The Crown Estate	TCE manages seabed use through impact assessment and issue of licences for activities. The Marine Data Exchange, established by TCE in 2013, was developed to store, manage and share offshore survey data collected by customers throughout the lifetime of their projects. https://www.thecrownestate.co.uk/en-gb/what-we-do/on-the-seabed/marine-planning/
Marine Recorder	Marine Recorder is a JNCC **benthic** survey data management system used widely within the UK's Statutory Nature Conservation Bodies to store and query benthic sample data across the UK's offshore and inshore waters. The system is able to store species occurrence data (with associated measurements), biotope information in the Marine Habitat Classification for Britain & Ireland and physical attribute data. The system maintains consistency and relationships between sample information, measurements and surveys allowing for accessible querying of the database. https://jncc.gov.uk/our-work/marine-recorder/
ESAS (historic) European Seabirds at Sea	ESAS data are managed by JNCC. Data collected by partners are supplied to JNCC who validate the data and collate it into the database using automated processes. Data access is by request to JNCC, who manually extract the requested data. Permission is given by consent from individual data owners, who require information on the individual or institute requesting how data will be used and what for, what the outputs of the work will, be, and whether the work is commercially driven. Where data owners are not content with their data being used, they can advise JNCC not to supply their data. Where data are requested for commercial projects, ESAS charge a fee of €0.1 per km^2 of effort, which is held by the ESAS partnership rather than going to those who collected the data. JNCC can also charge a nominal fee for the time taken in isolating and processing the data, where this is deemed necessary. This system is very labour intensive for JNCC as data managers. It also relies on each data provider nominating and maintaining a point of contact who can

	make judgements about data availability. The requirement to pay for data has put many potential users off accessing the dataset.
ESAS (current) European Seabirds at Sea	ESAS data owners are discussing passing management of the database to ICES. The details of this are still to be finalised, but could broadly be summarised as: • ICES keen to hold the data and will develop a web-based portal for data access that will sit among other ICES portals. ESAS partners responsible for data quality and QA, and data to be uploaded to database or harvested from partners databases annually, using automated processes. • Data products derived from the whole dataset will be open access (density surface at 10x10km squares, by month). ICES will develop and update the density surfaces. • Raw data available through the data portal. Open access data and restricted access data will be stored together. Open access data free to download. Requests for restricted access data generate an email requesting access to the data owner. • Data structure and metadata not dictated by ICES (apart from addition of columns to ensure coding consistencies across their databases – e.g. the addition of WoRMS codes). • Ownership of data fully retained.
MERP database/Bangor University Marine Ecosystems Research Programme	The MERP/BU database has the most up-to-date collation of data and was used for published analysis (Waggitt et al 2019, Evans and Waggitt 2019) and associated Atlas (in preparation). Users interested in accessing processed data from the database must first write to the original data providers and negotiate their own data sharing agreement. When permission has been granted, MERP/BU release processed sightings/effort from the database to the interested user. The user also has the option to negotiate the acquisition of additional/raw data from the original data provider, although this data may not compatible with that in the database. Under this arrangement, original data owners retain full ownership and control over their data.

6 References

Gilles, A., Scheidat, M., and Siebert, U., (2009). Seasonal distribution of harbour porpoises and possible interference of offshore wind farms in the German North Sea. Marine Ecology Progress Series 383: 295–307.

Hammond, P.S., Berggren, P., Benke, H., Borchers, D.L., Collet, A., Heide-Jørgensen, M.P., Heimlich, S., Hiby, A.R., Leopold, M.F., and Øien, N., 2002. Abundance of harbour porpoises and other cetaceans in the North Sea and adjacent waters. J. Appl. Ecol. 39, 361–376.

Hammond, P.S., Macleod, K., Berggren, P., Borchers, D.L., Burt, M.L., Cañadas, A., Desportes, G., Donovan, G.P., Gilles, A., Gillespie, D., Gordon, J., Hedley, S., Hiby, L., Kuklik, I., Leaper, R., Lehnert, K., Leopold, M., Lovell, P., Øien, N., Paxton, C., Ridoux, V., Rogan, E., Samarra, F., Scheidat, M., Sequeira, M., Siebert, U., Skov, H., Swift, R., Tasker, M.L., Teilmann, J., Van Canneyt, O, and Vázquez, J.A. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. Biological Conservation 164: 107-122.

Hiby, L (1999). The objective identification of duplicate sightings in aerial survey for porpoise. In: Garner, G.W. et al. (Eds.) Marine Mammal Survey and Assessment Methods, Balkema, Rotterdam. Pp 179-189.

Paxton, C.G.M., Scott-Hayward., L., Mackenzie., M., Rexstad. E., and Thomas, L. (2016) Revised Phase III Data Analysis of Joint Cetacean Protocol Data Resource JNCC Report No.517.

Tasker, M.L., Jones, P.H., Dixon, T, and Blake, B. F. 1984. Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardised approach. Nature Conservancy Council, 17 Rubislaw Terrace, Aberdeen AB1 1XE, Scotland. The Auk 101: 567-577.

Waggitt, J.J, Evans P.G.H. Andrade, J. Banks, A.N, Boisseau, O., Bolton, M., Bradbury, G., Brereton, T., Camphuysen, C.J., Durinck, J., Felce, T. , Fijn, R.C., Garcia-Baron, I., Garthe, S., Geelhoed, S.C.V , Gilles, A. , Goodall, M., Haelters, J., Hamilton, S., Hartny-Mills, L., Hodgins, N., James, K., Jessopp, M., Kavanagh, A.S., Leopold, M., Lohrengel, K., Louzao, M., Markones, N. , Martinez-Cediera, J., O'Cadhla, O., Perry, S.L., Pierce, G.J., Ridoux, V., Robinson, K.P., Santos, M.B., Saavedra, C., Skov, H, Stienen E.W.M., Sveegaard, S., Thompson, P., Vanermen, N., Wall, D., Webb, A., Wilson, J., Wanless, S., Hiddink J.G. 2019. Distribution maps of cetacean and seabird communities in the North-East Atlantic. Journal of Applied Ecology.

Annex II Workshop report

Joint Cetacean Data Programme (JCDP)

Stakeholder Workshop, Warrington 26/27 November 2019

Workshop Note

Working towards a data standard, data access agreement and host platform solution

Introduction and scene setting

The Joint Cetacean Data Programme (JCDP) aims to provide a growing, collated dataset of cetacean survey data to serve as a resource for stakeholders to inform conservation, research and policy needs.

Previous data collation exercises:

- JNCC published the **Atlas of Cetacean distribution in north-west European** waters
- The JNCC **Joint Cetacean Protocol** (JCP) developed from this collaboration, with the aim of standardising data for collated use, which enabled analyses such as the identification of persistent areas of use for harbour porpoise, and designation of Special Areas of Conservation (SAC)
- The NERC-funded **Marine Ecosystems Research Programme** (MERP) completed a further data collation exercise with a wider spatial and temporal scope, incorporating further datasets that were not available for JCP analyses.

Presentation: T*he Marine Ecosystems Research Programme, process and outputs.*

The MERP is the most recent collation of cetacean data. Over 2 million km of survey effort. Datasets have been analysed to predict monthly distribution maps for many cetacean species. However, funding has ended and data outside of MERP deliverables are not currently permitted for further use without first contacting data providers.

Standardising data

Task:

Spreadsheets were completed by data providers to establish the data fields that they currently use under their own data collection protocols. These are the basis for understanding where differences lie, and issues may arise in developing a standard approach to data collection. Data users discussed key uses of the JCDP.

Summary:

- Data providers are already collecting similar data in terms of fields recorded, enabling the identification of core fields across organisations.

- There was willingness to adapt terminology or method to a reasonable degree in order to support standardisation of the data.
- There is a need to clearly define each field to ensure consistency in how they are recorded and interpreted (e.g. behaviour categories, platform type and height, age classes and survey type).
- There are some fields which are not common across all types of survey e.g. aerial digital vs vessel observer which needs to be considered.

Distance sampling guidance: Natural England have developed a distance sampling guidance video in collaboration with CREEM. Potential to publish as part of the project on the webpage in support of the data collection standard. Two videos – short intro to principles, then a longer one on application to marine mammal surveys.

Next steps:

JNCC will produce an overarching spreadsheet of data fields with field definitions for the group to review. To be circulated to the PSG by 24th January 2020.

Scoping of a host platform

Presentation: *Overview of DASSH data archiving*

- DASSH can securely archive data at no extra cost to the provider, if data are supplied in the appropriate format (Darwin Core schema).
- Digitisation project for historic manual records
- Automated service to get data/metadata into other relevant platforms such as MEDIN, OBIS, EMODNet etc.
- Would need to develop a bespoke cetacean data guideline.
- Archive UK datasets only, although this may include some data outside of the UK EEZ. However, this may be an issue considering the aim of expanding the spatial remit of the JCDP beyond the UK.

Presentation: *ESAS and ICES*

- ESAS database previously held by JNCC, now migrating to ICES
- Desire to retain some level of control over data from some, others want open access
- Automated ingestion desirable to ease resource burden
- Strong support from ICES to host the data
- ESAS data available as a data product as agreed by the providers
- Underlying data can be requested via the portal, restricted will generate an email to the owner
- Data collated manually and sent to ICES, and also harvested from national datasets – requires input ‘somewhere’

Key points:

- Discussion is required on how to develop data products e.g. accounting for any differing opinions on modelling approach, spatial and temporal scales etc. There may

be issues with providing data products if based on more than one modelling approach. The approach needs to be consistent.
- Review needed at regular intervals to ensure products are fit for purpose and up to date in terms of methodology.
- Desire to retain involvement of a person with knowledge of cetacean data in the data flow to effectively manage it, including communication with providers and support in contributing data to the database. Build in a regular prompt to data providers to submit new data.
- Retain a communication role e.g. a JCDP data officer to maintain links with data providers and product outputs. Proactive data collation rather than a passive system will yield more up-to-date data. Also, to support building of validation/product tools for existing data providers and work with new providers to shape up data.
- Two stage validation – internal with the data provider to ensure quality control at point of collection and ensure it fits the standard/schema; JCDP data validation step.
- Need to retain the 'wider than UK' view to ensure any decisions can be extrapolated and are acceptable for EU/NE-Atlantic scale providers and users.
- DASSH may not be suitable for expanding beyond the UK but could play a role in archiving UK datasets.

Butterfly data model – case study: brings in UK funding for data collection and collation on a five-year basis, for use in support of policy assessments. Managed by JNCC. All data are open access. Funding pays for coordination costs.

One platform for data entry used by all data providers (which are limited in number compared to cetacean data). Possibly challenging to transfer to cetacean model across tens of groups in different fields. CEH produces processed products and runs trends with scripts developed over several years (open source). Scheme is annual with stats produced on trends and indicators. There's a small delay in data becoming available so data providers have a window in which to use data first.

Processed products – gov agencies have a 'shopping list' of outputs and provide funds to JNCC to manage. Source code is available so other schemes in other countries can mimic the same analyses. Usage of data is tracked to enable collaboration opportunities and feedback for providers to know how data are being used.

Need to (further) demonstrate value of the JCDP to policy to secure central funding.

Potential to publish the JCDP data standard and explanation of the science and justification as a paper.

Developing a data access agreement

Following the discussion at the workshop, the below outlines the main take-home messages for further consideration.

Open access *Raw data are available to download without restriction*	**Data categories:** Publicly funded data e.g. SCANS
Conditional access *Raw data are made available for production of agreed JCDP products but access beyond this may be subject to conditions*	**Data categories:** Non-Governmental Organisation (NGO) data – self-funded projects Industry/commercial data - funded as part of a plan or project Academic project data – discrete data collected for a specific academic project, not funded by public money

NGO data considerations

- Desire to maintain some control over how data are used to ensure it aligns with organisational values
- Raw data regularly submitted to JCDP for use in producing agreed data products for open access e.g. density surface models
- Access to raw data permitted for UK policy requirements e.g. statutory reporting
- Access to raw data permitted for industry use, but desirable to return a financial contribution from users
- Access to raw data for academic purposes to be by request
- Requirement for notification of how data are used to retain some control and enable collaboration opportunities

Industry/commercial data considerations

- Willingness to supply data for use
- Willingness to purchase data from the JCDP at a reasonable cost, if it saves resources elsewhere
- Sensitivities around access prior to consent
- Raw data regularly submitted to JCDP for use in producing agreed data products for open access e.g. density surface models
- Time-related restrictions possible for open access conditions e.g. post-consent
- Contractors don't own the data therefore access would need to be agreed with the data owner. Regulators could stipulate the need to provide data through the JCDP under agreed conditions e.g. time-related release.

Access options discussion

Option 1 – fully open access: not acceptable to NGO datasets, but acceptable for Gov-funded and possibly industry data.

Creative commons licence: provisions for free data use but must attribute the data to the provider(s). Can dictate for non-commercial use only.

Option 2 – data products (e.g. shapefiles): agreement across the group that this was an acceptable output of the JCDP, utilising raw data. Could be leverage for funding from industry to ease their data analysis burden. Need to be very clear on what products are required. Further products could be requested for a cost.

Option 3 – use-related: key option to develop the framework, using elements from other options.

Option 4 – time related: data made available after a specified time period – this is the model used by TCE. Probably most suitable for industry data in relation to consent. Post-consent, freely available? Could mean waiting for years before data are available which reduces ability to regularly update products.

Option 5 - tiered access: too complex, no longer a consideration.

Option 6 - tiered access, subscription or paid: This generates revenue for the project in order to enable e.g. further outputs to be produced or in support of maintaining the 'science and data coordinator' role. There would be a subscription, or single payment system for use of data, as agreed by stakeholders. Potential for project sponsorship also.

Income: any potential JCDP income – how is it used? Money going out to data providers is challenging to divide between all the cetacean data providers. How would it be split? Differences in how much it costs to collect different data. Differences in data collection methods and data standards currently.

If funding is secured for supporting data providers, where does funding support end? At data storage, processing and submission, or does it extend towards data collection also? Opens up the question re gov influence on data collection/access if funding is being provided to *collect* data. Influence on the training? Certified training courses in JCDP protocol?

Option 7 - access on request: JNCC desire to move away from this to realise the objectives of the JCDP. Elements may be integrated into other options. Data providers (mainly NGOs) want to know how data are being used.

If an element of the request remains, there will be a need for agreed response times (e.g. 72 hours) to requests in order to retain the JCDP reputation for all involved.

Need to bear in mind potential EU partners/contributors throughout to ensure they can be brought in when possible. Consider a few EU reps on the correspondence SG for JCDP (e.g. some from Germany and Netherlands already included). Share outputs of the workshop and where we are heading. Bring them in to the next round.

Actions: several actions were noted throughout the workshop which have all been addressed.

Post-meeting note:

A further meeting was held at the World Marine Mammal Conference in December 2019 as a result of Action 7. This meeting included NGO reps from SWF; WDC; HWDT; ORCA; MERP and NE, chaired by JNCC. The aim was to further discuss the perceived barriers to sharing data through the JCDP and consider a way forward that supports both the JCDP aims, and the needs of data providers.

Key points:

- There was reaffirmed agreement that open access to data products was acceptable
- Use of raw data would be granted for policy and statutory analyses conducted by the appropriate authorities
- A feedback protocol is required to enable notification of uses of data, so providers can inform members/volunteers etc. of how the data are being used
- Consideration required regarding potential commercial gain for users of the JCDP products/data and whether there is a need to request financial input in those cases. If so, where do the funds go?
- A JCDP MoU on authorship/acknowledgements is required so data users are clear on how work including JCDP data should be acknowledged
- We will need to manage capacity for data requests – currently little understanding of the potential volume of requests

Data access moving forward

There is still a desire to have an element of data request amongst NGO providers beyond the conditions listed in the first two bullets above, to enable management of how data are used as well as maximising collaboration potential. There was discussion of a 'soft start' in enabling some level of data request through the JCDP system, which may be relaxed as data providers become more comfortable with how the data are being used and how the system is functioning.

Next steps: JNCC to develop a draft data access agreement based on discussions.

Annex III Draft Data Standards Protocol

Joint Cetacean Data Programme

Draft Data Standards Protocol

Version 02

Guidance for JCDP data collectors

JNCC

Joint Cetacean Data Programme

Draft Data Standards Protocol

***This document is not a final version, but a record of progress following discussions to date.**

Contents

1 Overview 2
2 The need for a data standard 39
3 Spatial coverage 40
4 Data types 40
5 Data Tables: fields and vocabulary 41
6 Detailed guide to recording and submitting data 47
7 Frequency of data submission 50

1 Overview

The Joint Cetacean Data Programme (JCDP) is a portal for collation, storage and access of cetacean data collected at-sea via ship-based or aerial observer/digital methodologies. It is a growing resource, aiming to enable best use of all available data of comparable types from which to carry out analyses at relevant spatial and temporal scales to inform cetacean management, policy and conservation.

One of the key objectives of the JCDP is to work with data providers to synthesise the way in which data are collected and stored, in support of collating data into a central JCDP database. The JCDP Steering Group have agreed on a data standard under which to collect data that are compatible with the JCDP, to enable efficient submission to the JCDP. This document outlines that standard and data providers will need to work towards achieving the standard in order to result in compatible data with the JCDP. Please note, that this standard may have to be updated as the development of the database progresses.

2 The need for a data standard

For datasets to be collated into a single database, there needs to be a commonality at least between core fields within the data. Having a defined standard for data collection facilitates this. Standardising data offers advantages in four main areas: development of expertise and data quality, suitability of data for analyses, ease of data ingestion, and compatibility. These components also all contribute towards maximising the use of independent datasets.

Development of expertise and data quality

Where field data collection is standardised, surveyor expertise increases due to familiarity with an agreed, robust protocol. As a result, data quality improves through coordinated use of an appropriate survey methodology, as well as improved ability of observers to carry out accurate and effective surveys. Having an agreed standard will also support the development of new data collection initiatives, with a baseline from which to build a robust, compatible survey effort that can immediately contribute to the JCDP.

Suitability of data for analyses

Standardisation of data collection ensures that the requisite parameters are collected using the correct methods to do so. It also ensures that the data are recorded using the correct naming convention, taxonomy, and other associated coding. Recording supporting information e.g. spatial and environmental data, will also be controlled ensuring these data are suitable for use alongside other datasets.

Data ingestion process

Standardisation of datasets allows for automated upload and validation of data, saving time for both data owners and those hosting the data. The validation ensures that only those data that meet the standard will be stored, helping to maintain the quality of the data within the database.

Compatibility of datasets for combined use

Standardised data ease the burden of processing for those using the data. Data may need to be adapted to be used in analytical packages such as ‘R’ or ‘Distance’ but applying these adaptations from collection rather than to individual datasets further along the line, speeds up the process, eases data processing congestion and reduces opportunity for error. It also removes instances where data are completely incompatible for combined use, due to differences in data collection and storage methods.

The JCDP has appraised existing data standards and vocabularies and will aim to develop the JCDP standard so as to coordinate with those already established and adopted. As such, the data standard will be developed with input from relevant stakeholders such as MEDIN to become a recognised and appropriate standard to be applied widely.

3 Spatial coverage

The geographic range of JCDP datasets is initially focussed on UK data providers, although some datapoints may be outside the UK Exclusive Economic Zone (EEZ). The aim is to increase this geographical range to include the Northeast Atlantic, covering an area relevant to cetacean populations and movements to better inform management and conservation.

4 Data types

The data stored within the JCDP database include **effort-related** cetacean survey data collected via:

- Dedicated survey platform: ship-or aircraft observer survey
- Opportunistic survey platform: ship-based or aircraft observer survey
- Aerial digital imagery survey data

Although other types of data are not currently part of the JCDP, the JCDP dataset may be used in conjunction with others e.g. non-effort related observations; strandings or acoustic data to enable other analyses to be carried out.

5 Data Tables: fields and vocabulary

There are core data fields (e.g. species, locations) which are essential to enable effective use of the data for many applications and others which are not necessarily essential to answer key questions and complete primary analyses (e.g. observer name). The JCDP aims to coordinate the recording of core fields for cetacean data collection. The aim is not to dictate significant changes in how data are collected, particularly for established projects, but to provide a JCDP 'standard' to guide how data are recorded. The JCDP builds on the protocol established by its predecessor, the Joint Cetacean Protocol (JCP). There are three data tables: Identifiers; Effort and Environment; and Sightings Records. Where fields require further definition, this information is provided in Section 6. The agreed data fields are outlined below, with associated descriptions and definitions:

5.1 Identifiers

Field	Obligation	Drop down/ restricted format/ free text	Description	Drop down fields
DataOwner	Mandatory	FT	A free text field to capture the full organisation name that owns the data, as this may differ from whoever submits the data. This should be an organisation name where possible, written in full.	N/A
DataOwnerContact	Mandatory	RT	A restricted text field to record a contact for the data owner. A generic/organisation email is preferred over individuals due to GDPR, as well as continuity in retaining a current contact following personnel changes e.g. *info@company.co.uk*	N/A
DataSubmitter	Mandatory	FT	A free text field to capture the full organisation name that is submitting the data, as this may differ from whoever owns the data. This should be an organisation name where possible, written in full.	N/A
DataSubmitterContact	Mandatory	RT	A restricted text field to record a contact for the data owner. A generic/organisation email is preferred over individuals due to GDPR, as well as continuity in	N/A

			retaining a current contact following personnel changes e.g. *info@company.co.uk*	
PlatformType	Mandatory	DD	The type of survey platform from a drop-down menu - see Section 6.1 for platform definitions	Large ship, Med ship, Small ship, Aerial digi, Aerial obs
SurveyType	Mandatory	DD	The nature of the survey (dedicated or opportunistic)	Dedicated; Opportunistic
Methodology	Mandatory	DD	The methodology used (detail to be linked in organisational methodology metadata)	Single platform, line transect, distance sampling (SPLT:DS); Single platform, line transect, (SPLT); Double platform, line transect, distance sampling (DPLT:DS); Double platform, line transect (DPLT); Aerial digital (AD); Aerial observer (AO)
TargetTaxa	Mandatory	DD	Describes the focal taxa for the survey. Select drop down.	Marine mammals; Cetaceans; Megafauna, Seabirds & cetaceans, Seabirds and megafauna.

5.2 Effort & Environment

Field	Obligation	Drop down/ restricted format/ free text	Description	Unit	Drop down fields
StartDate	Mandatory	RF	dd:mm:yyyy start date of the survey	Date	N/A
EndDate	Mandatory	RF	dd:mm:yyyy end date of the survey	Date	N/A
StartTime	Mandatory	RF	hh:mm start time of survey effort (ship's time)	Time	N/A
EndTime	Mandatory	RF	hh:mm end time of survey effort (ship's time)	Time	N/A
SurveyID	Optional	FT	Restricted format field noting the unique identified assigned to the survey by the data owner. This is an optional field in support of retaining a reference between the data owner and the JCDP dataset.	Code	N/A
PlatformHeight	Mandatory	RF	Restricted format field with the height of observation platform. Maximum of 1 decimal place.	Metres	N/A
EffortType	Mandatory	DD	Nature of the survey effort waypoint recorded e.g weather change	Text (or code?)	On effort; Off effort; Change in environment; Change of course
PlatformSpeed	Mandatory	RF	Speed over ground	Km per hour	N/A
AngleOfSearch	Mandatory	DD	Search angle of survey team - if 'other', complete comments field. See Section 6.2.	Degrees	180 (forward); 100 (beam-port to 10); 100 (beam-starboard to 350); Other (complete comments field)
StartLatitude	Mandatory	RF	Start latitude of effort type.	Decimal degrees	N/A

StartLongitude	Mandatory	RF	Start longitude of effort type. Westerly longitudes will be prefixed with "-"	Decimal degrees	N/A
EndLatitude	Mandatory	RF	End latitude of effort type	Decimal degrees	N/A
EndLongitude	Mandatory	RF	End longitude of effort type. Westerly longitudes will be prefixed with "-"	Decimal degrees	N/A
SeaState	Mandatory	DD	Sea state using the Beaufort Scale. There is no requirement to submit data where sea state is over 6 on the scale as this is unlikely to be used in any analyses. https://www.metoffice.gov.uk/weather/guides/coast-and-sea/beaufort-scale	Number	1; 2; 3; 4; 5; 6+
SwellHeight	Optional	DD	Swell height in metres	Metres	0-1(Low); 1-2 (Med); 2-3 (High) 3+ (Very high)
Glare	Optional	DD	Sector of search area affected by glare, to the extent that it cannot be effectively searched	Percent	0-20%; 21-40%; 41-60%; 61-80%; 81-100%
Precipitation	Optional	DD	Precipitation that is affecting ability to search, from a drop-down list	Text	None; Rain; Snow; Hail;

					Fog
Visibility	Mandatory	DD	Visibility quality from platform to horizon.	Kilometres	0-1km; 1-5km; 5-10km; 10-15km; 15+
CloudCover	Optional	DD	Percent of sky above search area covered by cloud	Percent	0-20%; 21-40%; 41-60%; 61-80%; 81-100%
Turbidity	Optional	DD	Water clarity for aerial platforms		TBC
Comments	Optional	FT	Qualification of entries if required	N/A	N/A

5.3 Sightings Records

Field	**Field_type**	**Obligation**	**Drop down/ restricted format/ free text**	**Description**	**Unit**	**Drop down fields**
SpeciesScientific	Character	Mandatory	DD	Scientific name of species. Refer to WORMS (See Section 6.3)	Text	(List of species – see Section 6.3)
SpeciesCommon	Character	Mandatory	DD	Common name of the species	Text	List of species – see Section 6.3)

MinGroupSize	Numeric	Mandatory	RF	The minimum number of individuals of a single species in each sighting. (See Section 6.3)	Number	N/A
BestGroupSize	Numeric	Mandatory	RF	Best estimate of the number of individuals of a single species in each sighting. (See Section 6.3)	Number	N/A
MaxGroupSize	Numeric	Mandatory	RF	Estimate of the maximum number of individuals of a single species in each sighting. (See Section 6.3)	Number	N/A
NumberOfCalves	Numeric	Mandatory	RF	Proportion of calves in the group as a number	Number	N/A
RadialDistance	Numeric	Optional	RF	Distance of animal(s) in metres from the observer. This might be recorded by eye with or without a rangefinder or converted from a reticule binocular measurement in the field.	Metres	N/A
RadialAngle	Numeric	Optional	RF	Radial angle of sighting, where 0° is directly ahead on the platform track. Should be recorded with use of an angleboard or ship's compass in the field.	Degrees	N/A
Confidence_ID	Character	Mandatory	DD	Confidence in species id from a drop-down list - See Section 6.3	Text	Definite; Probable; Possible.
Behaviour	Character	Optional	DD	Primary behavioural information of note from a drop-down list. Further information can be recorded in the comments field if required.	Text	Swimming/Travelling; Breaching; Feeding; Logging/Resting; Vessel Avoidance; Spy Hopping.

6 Detailed guide to recording and submitting data

Below is a detailed description of how fields should be recorded:

6.1 Identifiers – detailed description

PlatformType: The type of survey platform from a specified drop-down menu. Vessel size is defined as vessel length in metres:

Category	Drop down code	Definition
Large ship	LargeShip (>120m)	Vessels with a length of 120m or more
Medium ship	MedShip (80-120m)	Vessels with a length of between 80m and 120m
Small to Medium ship	SmallMedShip (25-80m)	Vessels with a length of between 25m – 80m
Small ship	SmallShip (<25m)	Vessels with a length of 25m or less
Aerial digital	AerialDig	Aerial survey conducted via digital photography
Aerial observer	AerialObs	Aerial survey conducted with an onboard observer

SamplingType: The nature of the survey from a specified drop-down menu. This is to enable filtering of results for dedicated cetacean survey, from other survey types.

Category	Drop down code	Definition
Dedicated	Dedicated	A survey event planned and dedicated to cetacean survey (e.g. SCANS) This may also include other taxa as identified in the *TargetTaxa* field.
Opportunistic	Opportunistic	A survey event that takes opportunity of a survey/platform with other objectives (e.g. ferry survey; survey where cetacean records are opportunistic)

Methodology: The methodology used to collect the data from a specified drop-down list. Individual organisational metadata within each survey methodology should be linked, containing specific details on how the methodology is applied.

Category	Drop down code	Definition
Single platform line-transect, distance sampling method	SPLT:DS	A line transect survey where the observer (team) operates from one platform on the ship and distance sampling methodology is applied.
Single platform line-transect	SPLT	A line transect survey where the observer (team) operates from one platform on the ship, but distance sampling methodology is ***not*** applied.

Double platform line-transect, distance sampling method	DPLT:DS	A line transect survey where there are two observer teams operating from two independent locations with the same outlook, on a single platform, and distance sampling methodology is applied.
Double platform line-transect	DPLT	A line transect survey where there are two observer teams operating from two independent locations with the same outlook, on a single platform, and distance sampling methodology is ***not*** applied.
Aerial digital	AD	A survey taking place by aircraft using digital imagery (stills or video) to record sightings and sampling effort
Aerial observer	AO	A survey taking place by aircraft using onboard observers to record sightings and sampling effort

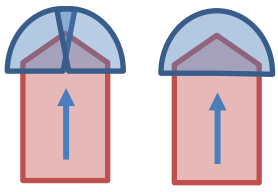
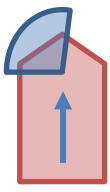
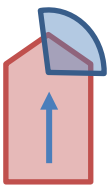
TargetTaxa: Describes the focal taxa for the survey from a specified drop-down list. This captures whether observers were solely searching/recording cetaceans, or if it was a combined survey or cetaceans were not a specific focus, which may impact the detection function of analyses.

Drop-down selection:

- Marine mammals
- Cetaceans
- Megafauna
- Seabirds & cetaceans
- Seabirds & megafauna
- Seabirds

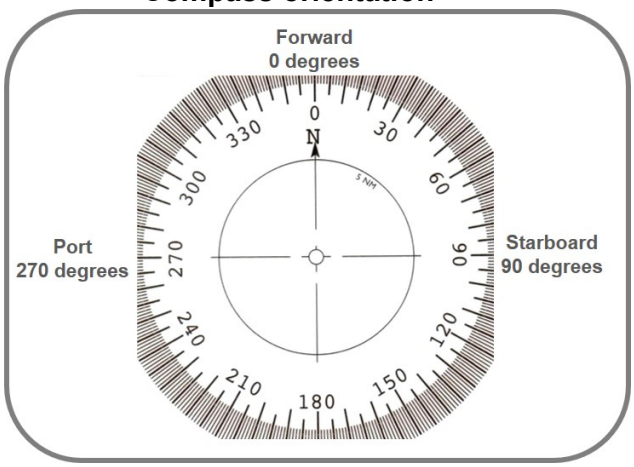
6.2 Effort & Environment – detailed description

AngleOfObservation: Search angle of observer in degrees, from a specified drop-down list. If 'other', the comments field should be completed.

180 (forward)	100 (beam-port to 10)	100 (beam-starboard to 350)	Other (complete comments field)
			For other variations on the specified angles of search. Descriptive format should follow the same style as the drop-down as much as possible.

All angles should be taken with 0 degrees at the direction of travel.

Compass orientation



6.3 Sightings Records – detailed description

Species: drop-down list of scientific names for cetacean species as recorded in the World Registry of Marine Species (WORMS) http://www.marinespecies.org/

There are also categories for unidentified sightings:

Unidentified cetacean sp	For records where it is not possible to distinguish what group of cetaceans this sighting belongs to, but there is a level of confidence that an animal was sighted (confidence level still to be recorded).
Unidentified baleen whale sp	For records of unidentified animals, that can be confidently identified as a species of baleen whale (confidence level still to be recorded).
Unidentified toothed whale sp	For records of unidentified animals, that can be confidently identified as a species of toothed whale (confidence level still to be recorded).
Unidentified whale sp	For records of unidentified animals, that can be confidently identified as a species of whale (confidence level still to be recorded).
Unidentified dolphin/porpoise sp	For records of unidentified animals, that can be confidently identified as a species from the dolphin/porpoise family (confidence level still to be recorded).

To note:,the dataset is of higher quality when records are given to the best level of confidence. Therefore, where a record is 'definitely' a whale, this is more useful to record than a 'possible' named species. Please add relevant notes to the comments field in support of this record, including notes on potential species and confidence in that judgement.

MinGroupSize: Restricted format field - numerical value recording the estimate of the smallest possible number of animals observed in a sighting. This is to essentially to act as a basic coefficient of variation (CV) to illustrate confidence in the value noted as the sighting group size.

BestGroupSize: Restricted format field - numerical value recording the best estimate of the actual number of a species in a sighting. This may be the same as the min and max values where observers are highly confident in the number count.

MaxGroupSize: Estimate of the largest possible number of animals in a sighting. As with the MinGroupSize, this is to essentially result in a coefficient of variation (CV) to record confidence in the value noted as the sighting group size.

Confidence_ID: From a drop-down list, the observer confidence in the species id:

Definite	The observer is confident in the species Id following presence of several cues that confirm identification. Also to be recorded for unidentified records, in terms of confidence in the selected category.
Probable	The observer is not completely confident in the species id, but presence of cues and features enables a moderate level of confidence in the species id. Also to be recorded for unidentified records, in terms of confidence in the selected category.
Possible	The observer is not confident in the species id or even presence of an animal, as the cue was unclear. Also to be recorded for unidentified records, in terms of confidence in the selected category.

6.4 Summary of data submission requirements

- Data submission will be made through the JCDP portal
- Ahead of submission, all data should be prepared according to the JCDP standard and saved in csv format with separate effort & environment table and sightings table
- Casual, unsystematic, observations not associated with formal commencement of search effort must NOT be included
- Personal data should not be submitted. Information on individual observers should stay with the data owner.

Data collector detailed survey protocols and associated information

The JCDP Data Standards Protocol is designed to coordinate and create consistency amongst cetacean data collectors, identifying core fields and defining how they should be recorded. It is not a comprehensive protocol to replace existing or future data collection protocols of individual collectors. Details should be supplied of the specific survey protocol, to be made available alongside any data published through the JCDP.

7 Frequency of data submission

Data collectors are able to submit data at any point throughout a year. However, a data call will be initiated annually in order to ensure the database remains up to date for the production of the data products (as agreed buy the steering group).

The data call will likely be initiated at the end of the primary data collection season in late Autumn, with the aim of producing the annual data products in Spring each year, incorporating any new data.

Data that are collected on an ad hoc basis or forming part of short-term projects can be submitted to the JCDP as and when relevant, with a view to ensure data can be included in production of annual products in Spring where possible.

Annex IV Data validation evaluation

This annex summarises and compares validation processes for 3 existing collations of at-sea surveys: Marine Ecosystems Research Programme (MERP), Joint Cetacean Protocol (JCP) and European Seabirds at Sea (ESAS). Those collecting and providing data are referred to as suppliers whereas those collating data are referred to as collators. Manual validation describes checks achieved by human inspection of graphs, maps and spreadsheets. Automated validation describes checks achieved by computer algorithms.

1: Marine Ecosystems Research Programme (MERP)

The following summary of the validation process performed by Bangor University/Sea Watch Foundation is based on the published paper (Waggitt et al. 2019). As the authors performed the validation, the summary is more detailed than for the JCP and ESAS.

1.1 Preparation

Suppliers provide data in various formats. Collators then import and amalgamate these data in R. Manual validation focussed on essential information (date/time, coordinates, sea state) using plots and tables. Erroneous data were removed if ambiguous or corrected if obvious. Essential information not included in data was sourced from reports/papers/suppliers and added at this stage. The effort spreadsheet for automated validation was presented as a list of sections (start/end coordinates and times) and essential information (platform, method, sea state, observer height, supplier). The sightings spreadsheet was presented as a list of sightings (species, counts, coordinates and times) and essential information (distance/bearings, platform, method, sea state, supplier, observer height and supplier).

1.2 Effort

The following describes the automated validation process for effort data:

1. Duplicate rows showing identical information were omitted.
2. The removal of invalid speeds. This is based on calculating the mean speed for the platform conducting the survey and omitting sections that occurred beyond an acceptable range from this mean. The calculation of the mean speed excluded extreme outliers from particularly erroneous data, which could have skewed the resultant value of mean speed. For vessel surveys, where slow speeds are possible, the acceptable range was up to the mean ± (mean/2). For aerial surveys, where slow speeds are not possible, the acceptable range was the mean ± (mean/4). These thresholds were based on visual inspection of data across suppliers.
3. The removal of data on land. This was achieved by constructing a 1km x 1km raster layer whereby cells with values of 1 showed landmasses, and those with values of 0 showed sea-surfaces. It was assumed that platforms travelled on a consistent bearing within each section. Therefore, section routes were represented by a series of points at ~1km intervals between the start and end coordinates. If any point occurred within a cell with a value of 1, then it was omitted from the data.
4. A series of summary maps and figures were produced. These maps and figures were checked with associated reports/papers/websites for a “sense-check”. In some cases, suppliers were contacted for additional details on survey routes and timings.

1.3 Sightings

The following describes the automated validation process for sightings data:

1. The removal of sightings without associated effort by checking whether sightings occurred within the timespan (earliest and latest time) of validated sections.
2. The removal of sightings on land. This was achieved by constructing a 1km x 1km raster layer whereby cells with values of 1 showed landmasses, and those with values

of 0 showed sea-surfaces. If any sighting occurred within a cell with a value of 1, then it was omitted from the data.

3. A series of summary maps and figures were produced, and checked with associated reports/papers/websites for a "sense-check". They were also assessed subjectively using knowledge of species ecology and distribution, and collator awareness of sightings in the corresponding region, season and year. The latter were only used to identify potential problems with code and validation, rather than to omit data simply because it did not confer with what the authors expected or believed. In some cases, suppliers were contacted for additional details on to query rare or atypical sightings. However, data were unadjusted and not queried wherever possible.

1.4 Evaluation

Positives

1. Manual validation by collators reduces staff-time and costs for suppliers.
2. "Sense-checking" by collators, based on personal communications with suppliers and consultation of literature, retains some realism and connectivity with the data.
3. Omission of erroneous speeds is tailored to platforms/suppliers (considering variation in methods and operation), improving detection of erroneous data.
4. Consideration of landmasses omits non-obvious erroneous data near coastlines.

Negatives

1. Manual validation and "sense-checking" increases staff-time and costs for collators.
2. Because manual validation is not performed by those conducting surveys, simple errors most probably culminate in data removal rather than correction. For example, effort in coastal areas may be wrongly omitted if vessels perform tortuous routes around complex topography and do not record coordinates regularly.
3. No automated validation for removing sightings with coordinates that are too far away from the transect line to be realistic (i.e. 10-100s of km). In most cases, these would have been detected in manual validation, but this cannot be guaranteed entirely. However, for the Waggitt et al (2019) analyses, these sightings were automatically removed when gridding effort and sightings data for Species Distribution Models.

2: Joint Cetacean Protocol (JCP)

The following summary of the validation process performed by CREEM in the JCP analyses was based on the published report (Paxton et al. 2017) and an R code provided by the lead author (Charles Paxton, personal correspondence). As this is based on an informed assessment and interpretation of the code, the summary is less detailed than for MERP.

2.1 Preparation

Each provider appears to have provided a spreadsheet, as the collator imports spreadsheets into R before automated validation. The automated validation process in R neither addresses obvious errors nor typographic errors, suggesting that major issues were corrected in manual validation by suppliers. The effort spreadsheet for automated validation was as a series of waypoints (coordinates and times) with essential information (platform, method, sea state, observer height, supplier). The sightings spreadsheet for automated validation was presented as a list of sightings (species, counts, coordinates and time) and associated information (distance/bearings, platform, method, sea state, supplier, observer height and supplier).

2.2. Effort

The following describes the automated validation process for effort data:

1. The removal of any invalid dates or times. Warnings are also provided if dates and times are not presented in chronological order for a certain vessel. There also appears to be an automated correction to convert all times into GMT.
2. A check that the number of start points matches the number of end points. These checks are presumably because effort is stored as a track showing a sequence of coordinates (1 per row), rather than sections with start and end coordinates (2 per row). If it is the former, this explains the need to check whether date and times are presented in chronological order (see stage 1), as an incorrect chronology in this format would cause erroneous distances travelled and speed of travel.
3. A check that the start and end points are in the correct number for each platform. This process assumingly double-checks that stages 1 and 2 have worked.
4. The removal of invalid speeds. This is based upon the maximum speed for the platform conducting the survey. It seems that this approach simply removes sections with speeds exceeding the maximum speed possible for the platform in question. It is unknown from where the maximum speed is sourced, i.e. whether information is provided by the supplier, or a realistic maximum speed is estimated. It is assumed that the later scenario is more likely, given the large number of platforms and extensive time period i.e. obtaining details for all vessels in JCP seems unlikely.

2.3 Sightings

The following describes the automated validation process for sightings data:

- The removal of sightings without associated effort by checking whether sightings occurred within the time-limits (earliest and latest) of validated sections.

2.4 Other Comments

There was no automated validation for effort and sightings on landmasses. However, there were manual validations for effort and sightings that very-clearly on landmasses or in incorrect locations. Extensive “sense-checks” are not mentioned. Therefore, subtle-errors or errors only obvious to those “in-the-know” could remain within the database following automated and manual validation (i.e. species in atypical areas, effort in incorrect times/areas).

2.5 Evaluation

Positives

1. Emphasis on automated validation decreases staff time and costs for collator.
2. Because manual validation is performed by those performing surveys, simple errors most probably culminate in data correction rather than removal.
3. Thorough checks for date/time reduce likelihood of error.

Negatives

1. The need for providers to perform manual validation increases staff-time and costs.
2. The apparent absence of “sense-checks” means that data considered valid during validation but clearly incorrect to those ‘in the know’ is retained.
3. Use of maximum theoretical speed rather than average recorded speed of platforms may retain erroneous data in vessels that operate at particularly slow speeds.
4. No automated validation for removing sightings with coordinates that are too far away from the transect line to be realistic (i.e. 10-100s of km). In most cases, these would have been detected in the manual validation, but this cannot be guaranteed.
5. No automated validation for effort and sightings on landmasses.

3: European Seabirds At-Sea (ESAS)

The following summary of the validation process for ESAS is based on JNCC documents (Mark Lewis, personal correspondence). As this is based on an informed assessment and interpretation of the document, the summary is less detailed than for MERP.

3.1 Preparation

Each supplier appears to send spreadsheets to the collator. The collator then converts the original data into the coded data required for amalgamation.

3.2. Effort and Sightings

Whilst automated validation was performed, detailed information on this process was unavailable at the time of writing. However, automated validation focussed on identifying missing-information and ranking of quality based upon segment duration i.e. increasing time between waypoints would indicate missing/inaccurate information. Potentially erroneous points were flagged for inspection during manual validation rather than being removed. Manual validation then included inspection of plots and maps for incorrect coordinates, unrealistic speeds, data on landmasses, unusual coordinates and discrepancies between effort/sightings. The decision to remove or retain data seems based on a qualitative judgement. The spreadsheet resulting from manual validation is a series of waypoints (times and coordinates) with associated information (platform, method, sea state, observer height, supplier). Sightings and effort are included on the same spreadsheet. Effort data are duplicated when several sightings occur at the same time and location. For each sighting, appropriate information is included as additional columns (species, count, distance).

3.3 Other Comments

Extensive “sense-checks” are not mentioned. Therefore, subtle-errors or errors only obvious to those “in-the-know” could remain within the database following automated and manual validation (i.e. species in atypical areas, effort in incorrect times/areas).

3.4 Evaluation

Positives

1. Manual validation by collators reduces staff-time and costs for providers.
2. Qualitative judgement on potentially erroneous data reduces incorrect removal.

Negatives

1. The need for collators to perform manual validation and qualitatively assess potentially erroneous data increases staff-time and costs.
2. The apparent absence of “sense-checks” means that data considered valid during validation but clearly incorrect to those ‘in the know’ are retained.

Annex V Draft Data Policy

Joint Cetacean Data Programme

Draft Data Provision, Access and Use Policy

Joint Cetacean Data Programme

Draft Data Provision, Access and Use Policy

***This document is not a final version, but a record of progress following discussions to date.**

1. Overview 57
2. Provision of data to the JCDP 57
3. Access to products and raw data 58
   3.1 Use of the raw JCDP dataset for UK policy 58
   3.2 Agreed uses of JCDP raw data by statutory organisations. 59
   3.3 Conditional access: restrictions on data access 60
4. Personal data 60
5. Data requests 61
   5.1 Notification of project outputs 62
6. Using JCDP data beyond the scope of the submitted project outline 62
7. Collaboration 62
8. Acknowledgements and authorship 62

List of Tables

Table 1: Data access overview 58
Table 2: UK organisations with permissions to use the JCDP dataset for outputs (as outlined in Table 3) 59
Table 3: Agreed data uses for statutory purposes 59

List of Figures

Figure 1: Data request process 61

1. Overview

The Joint Cetacean Data Programme (JCDP) is a portal for collation, storage and access of cetacean data collected at-sea via ship-based or aerial observer/digital methodologies. It is a growing resource to enable best use of all available data of comparable types from which to carry out analyses at relevant spatial and temporal scales to inform cetacean management, policy and conservation.

This document sets out the responsibilities that the data owners/providers and end users of the data have with respect to making data contributions to or requesting data from the JCDP. To summarise:

JCDP obligations of **Data Owners/Providers**

- By submitting data to the JCDP, you support the ethos of the JCDP which is to improve and facilitate sharing and accessibility of data to underpin activities for the purpose of cetacean management, policy and conservation.
- All data submitted to the JCDP should be compliant with the JCDP Data Standards Protocol.
- All data submitted to the JCDP will be available through two levels of access: i) instant access or ii) access by request.
- Personal data should not be submitted to the JCDP but remain with the data owner.

JCDP obligations for **End Users**

- Data downloaded/requested for use through the JCDP will only be used expressly for the purpose stated in the data request.
- Data will not be shared for use by, or sold to Third Parties after download, and will not be used beyond the original purpose stated.
- Any suspected errors in the downloaded data should be reported to the JCDP coordinator.
- Acknowledgement of the JCDP and data owners/providers must be applied according to the guidance within this document.
- References to the outputs of the JCDP data use should be provided to the JCDP on completion of the project.

This policy will be reviewed annually and updated as and when required in consultation with JCDP stakeholders.

2. Provision of data to the JCDP

By submitting data to the JCDP, there is an understanding that data owners/providers support the ethos of the JCDP which is to improve and facilitate sharing and accessibility of data to underpin activities for the purpose of cetacean management, policy and conservation. When providing data and associated meta data to the JCDP, it is understood that:

a) Provision of data is subject to the conditions outlined in this document.
b) Although there may be some level of quality control and validation during submission to the JCDP, quality assurance of data is the responsibility of the data owner.

3. Access to products and raw data

Agreed data products produced using the JCDP dataset are openly accessible through the JCDP portal. Download of the data is instant unless stated otherwise, in which case a notification will be generated to the data owner(s) following submission of a download request (see Section 5: Data requests).

The JCDP database holds data from multiple providers collected through a variety of means as outlined in the JCDP Data Standards Protocol, with two levels of access: Instant access and Access by request (see Table).

Table 1 Data access overview

Access type	Data	Description
Instant access	Data products	Data products using the JCDP dataset produced annually as agreed by the JCDP Steering Group, made freely available for use. [list products once agreed]
	Raw data available to download without restriction	Publicly funded data e.g. the Small Cetacean Abundance in the North Sea and Adjacent waters (SCANS) survey data, and data from other providers who have granted instant access
Access by request	Raw data available for request through the JCDP	Some data within the JCDP require a communication to the data owner in order to request access, as outlined in Section 4. An exception exists in order to facilitate rapid response to assessment needs within statutory processes, whereby raw data can be used by the organisations listed in Table 2 for the production of outputs listed in Table 3, as agreed by the JCDP Steering Group.

The data products will be reviewed annually in the first three years of the project, and then every two years going forward with input from the established Steering Group to ensure they remain up-to-date and fit for purpose.

3.1 Use of the raw JCDP dataset for UK policy

All raw data submitted to the JCDP will be openly available for UK statutory uses (see Section 3.2) by organisations with an advisory capacity to UK Governments (listed below in Table) as agreed by the JCDP Steering Group. Data will need to be downloaded for each new use, and not permanently stored by these users (see Section 6: Storing JCDP Data). Data downloads will be recorded as with all data requests and the data owners will be notified in order to maintain tracking of data uses. The JCDP coordinator (if appointed) will

track and maintain a list of data uses from the download alerts, to publicise in the JCDP portal.

Table 2: UK organisations with permissions to use the JCDP dataset for outputs (as outlined in Table 3)

1	Joint Nature Conservation Committee (JNCC)
2	Scottish Natural Heritage (SNH)
3	Marine Scotland Science (MSS)
4	Natural Resources Wales (NRW)
5	Natural England (NE)
6	Department of Agriculture. Environment and Rural Affairs (DAERA) – in the Governmental advisory capacity.

3.2 Agreed uses of JCDP raw data by statutory organisations.

One of the key aims of the JCDP is to enable continuity and accessibility of best available evidence on which to base policy and management decisions. The data products which will be produced regularly to help fulfil some of these needs. In cases where further analyses are required, raw data provided to the JCDP will be made available to the organisations listed in Table for uses defined in

Table , in order to:

1. enable rapid and relevant analyses to meet the UK's assessment and reporting obligations under national and international legislation;
2. support development of advice prepared to inform Government policy and decisions.

If the raw data were required for uses not listed in Table 3, a request would need to be made to the relevant data owners.

Table 3: Agreed data uses for statutory purposes

	Driver	**Output(s)**
1	OSPAR	Assessment and reporting on cetacean indicators, including Quality Status Reports; Identification of Marine Protected Areas
2	Marine Strategy Framework Directive (MSFD)/UK Marine Strategy	Assessment and reporting on UK cetacean indicators; Identification of Marine Protected Areas
3	EU Habitats Directive/Bern Convention	Assessment and reporting of cetacean conservation status;

4	Agreement on the Conservation of Small Cetaceans of the Baltic, North East Atlantic, Irish and North Seas (ASCOBANS)	Contribution to national reporting to the Advisory Committee.
5	UK Dolphin & Porpoise Conservation Strategy	Contribution to assessment of status. Annual and seasonal distribution and abundance in support of actions within the Strategy.

These listed uses will be reviewed annually in the first three years of the project, and then every two years going forward with input from the established Steering Group in consultation with the Steering Group.

3.3 Conditional access: restrictions on data access

The primary objective of the JCDP is to facilitate the collation and subsequent accessibility of cetacean monitoring data, and it is expected that data provided to the JCDP is in support of this goal. However, some data submitted to the JCDP will have restrictions applied and in the event of a download request, a communication will be generated to request permission from the owner(s) before download is enabled. These restrictions will likely be needed for commercially sensitive data but data owners also have the right to refuse use of data if e.g. there are aspects of an intended use that are contrary to the organisational values. Any decline of requested data will result in a report generated to the JCDP detailing the reasons. Any reports that are received will be used as a feedback method to identify whether access of data through the JCDP is fair and any requests declined are justified and do not run contrary to the ethos of the project. If there appears to be issues surrounding the release of data, the steering group will be consulted to consider a way forward in improving data owner confidence in enabling use of data.

If there are suspected errors in a dataset downloaded from the JCDP, the JCDP coordinator should be contacted with notification of the suspected errors so this can be followed up with the data owner(s).

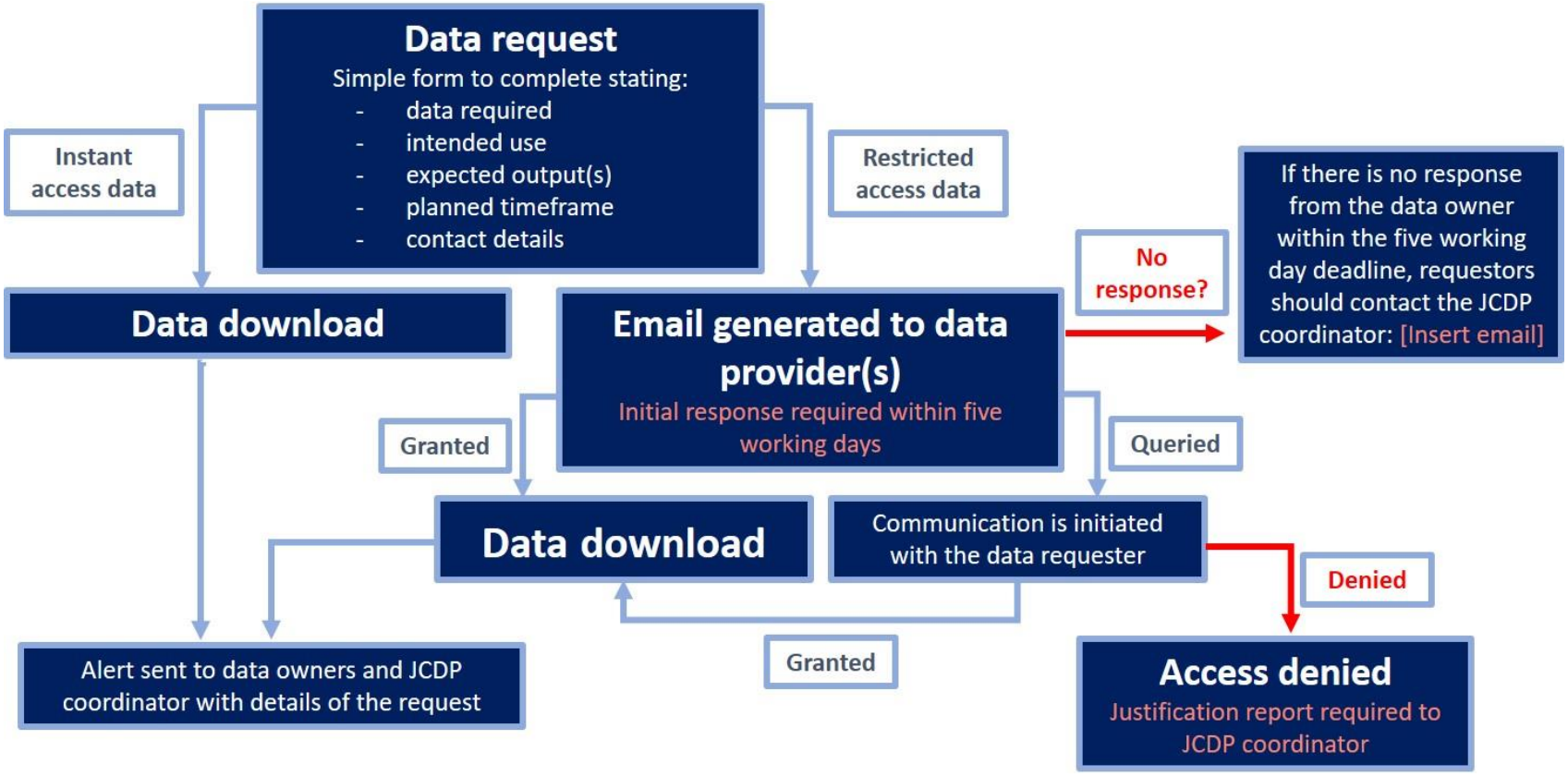
4. Personal data

Personal data should not be submitted to the JCDP wherever possible, but remain with the data owner e.g. observer names, individual email contacts. Owner/provider-specific data should be in the form of organisational email contacts and depersonalised records. However, if personal data *are* submitted to the JCDP for any reason, they will be stored in compliance with the General Data Protection Regulation (2018) and will not be provided as part of a JCDP data download. If data users have a need for personal data, data owners should be contacted directly.

5. Data requests

Raw data can be requested from the JCDP through the portal using a variety of queries e.g. by species, spatial extent; temporal parameters etc. Alerts will be returned to the JCDP mailbox, for each data request in order to track usage. Instant access data will be accessible for download immediately, whereas datasets requiring permissions from the data owner(s) will be flagged, and an email generated to the relevant data owner(s). Responses to a data request are expected from the data owner within 5 working days of sending, although this may simply be an initial confirmation of receipt of the request in the first instance. For restricted access, only the data owner can give permission for download of the data.

Figure 1: Data request process



5.1 Notification of project outputs

In order to measure the effectiveness of the JCDP, there is a request for projects using the JCDP data to communicate any outputs associated with that data use. As a result, the JCDP will not only be able to support communication of these outputs, but also track the benefits of the resource to encourage participation and use. It is also of interest and benefit for data owners/providers to know how the data are being used so this can be communicated to funders/members and interested parties, to ensure the data collection is recognised as having an important contribution to the growing database.

Project outputs should be communicated to the JCDP coordinator via the email contact [insert] before or on completion of the project. Project outputs may be listed/linked through the JCDP webspace if agreed with the project manager.

6. Using JCDP data beyond the scope of the submitted project outline

Data downloaded from the JCDP through the request route, **must not** be used following completion of the project analyses described in the request for data. Data also must not be shared for use by Third Parties. If further uses are identified beyond the original project scope, a subsequent data download request should be made through the JCDP with an outline of the intended use. This is to:

a) Enable the JCDP and associated data providers to track use and, therefore, value of the JCDP;
b) Ensure datasets are up to date and all data used have permissions granted;
c) Support opportunity for collaboration in using the data to best effect (see Section 8)).

A detailed terms of use contract will accompany any data download, which will outline the use, storage and sharing of data downloaded from the JCDP. This will be developed as the JCDP project develops.

7. Collaboration

The JCDP aims to encourage collaboration in using the data, both with the providers of a particular dataset, and/or others involved in the JCDP who are able to provide expertise to strengthen any outputs. In order to help facilitate this, all data downloads generate a notification to relevant JCDP data providers and the JCDP coordinator [if appointed] to enable communication between relevant parties. There is no *requirement* to collaborate when accessing JCDP data, but it is encouraged to ensure the data are used correctly with input from those who best understand the data.

8. Acknowledgements and authorship

When using JCDP data, there is a requirement to provide appropriate acknowledgement to the project and data owners in order to recognise the contributions of those inputting data to the resource and the subsequent benefits to others.

If any data are sourced through JCDP, then acknowledgement of the programme and relevant data owners/providers must be given:

In order to identify the appropriate acknowledgment, the following guidance applies:

i. If using the entire JCDP dataset, or it is the only source of data for a piece of work, acknowledgement should include the JCDP standard acknowledgment format:

JCDP standard acknowledgement format:

This work was completed with data provided by the Joint Cetacean Data Programme and associated data providers https://jncc.gov.uk/our-work/joint-cetacean-data-programme/. [Link will be the most relevant link to list of data providers]

ii. If using a subset of the JCDP dataset the following guidance applies:

The JCDP standard acknowledgement format (above) should be in the acknowledgements. Furthermore, there is a requirement to appropriately acknowledge the primary data owners/providers contributing data to the project. Where the dataset comprises data from 5 or fewer unique data owners and the contribution from an individual data owner comprises a significant proportion of the dataset, the data owner(s) should be offered an individual acknowledgment.

Data owners collaborating on projects as a result of JCDP data access, either through provision of a significant portion (>20%) of the data used, and/or support in carrying out the project, are free to discuss co-authorship with the project lead, to be agreed in the initial stages of the project.